Productivity and Labor Allocation Within Teams of Knowledge Workers*

Ruo Shangguan
Jinan University

Jed DeVaro
California State University, East Bay

Hideo Owan
RIETI, Waseda University

June 16, 2023

Abstract

We study the optimal functioning and productivity of teams of white-collar knowledge workers, using data from a large Japanese architectural and engineering consultancy firm. Most team working hours are supplied by only a few workers, evoking the Pareto Principle of business management. A new model of within-team labor allocation replicates the distributions of team productivity and within-team labor allocation and permits decomposition of a nearly 7.5% team productivity increase (resulting from the 2008-2009 financial crisis) into parts due to increases in individual worker productivity (2.5%) and within-team labor reallocation (5%).

Keywords: team production, labor productivity, working hours, allocation of labor, labor changes during recessions, Great Recession, global financial crisis

JEL classification: J23, J24, M50, M54

*This study is conducted as a part of the Project "Productivity Effects of HRM Policies and Changing Employment Systems" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). A previous version of this study, titled "Enhancing Team Productivity through Shorter Working Hours: Evidence from the Great Recession", appears as a discussion paper on the RIETI website. Helpful comments from the editor (Iourii Manovskii), three anonymous reviewers, Sachiko Kuroda, John Pencavel, Guangyu Pei, Kathryn Shaw, Katsuya Takii, Kensuke Teshima, Yuta Toyama, Daniel Xu and seminar participants at the 2020 Organizational Economics Workshop (Sydney, Australia), Waseda University, Hitotsubashi University, Society of Labor Economists 2021 conference, JEA 2021 Spring meeting, AASLE 2021 conference, and Jinan University are gratefully acknowledged.

1 Introduction

Teamwork is an engrained feature of the modern workplace[1] and has increased in complexity over time, as the nature of work has become more global, virtual, project focused, and enmeshed in new technologies. Understanding how to manage teams for optimal productivity in increasingly complex environments is essential for business managers. A central question managers face in that regard is how to assign labor hours across workers within teams, to maximize team productivity. If an increase in demand for the firm's product or service necessitates adding more worker hours, which team member(s) should work more? In the reverse case of a negative demand shock, which team member(s) should work less? What are the implications of these decisions for team productivity?

Daunting data demands have hampered academicians' ability to provide much guidance or insight into these issues, particularly for white-collar teams of knowledge workers. Much of the existing empirical work on teams and productivity focuses on lower-skilled teams. The dearth of empirical research on within-team labor allocation for knowledge workers leaves unanswered, for example, the question of how important "stars" are as drivers of team productivity. A wealth of anecdotal information suggests that stars effectively carry their teams.[2] That widely recognized idea is encapsulated in the Pareto Principle of business management, which states that 80% of the work that a team accomplishes is completed by only 20% of the team's membership.[3] If a team has just a handful of members, as many do, the Pareto Principle implies that often a single worker is responsible for most of the team's output. Anecdotal evidence notwithstanding, the truth or falsity of the Pareto Principle as it applies to teams of knowledge workers remains an open empirical question.

This study empirically investigates the determinants of team productivity in construction project design teams, exploiting unique data on construction projects in a Japanese architectural and engineering consultancy firm during the years 2004 to 2016.[4] That times-

[1]Using a cross section that is representative of U.K. establishments in 1998, Table 6 of DeVaro (2006) documents that the average (across establishments) proportion of the establishment's largest occupational group that works in formally designated teams is 0.77. That statistic remained relatively stable more than a decade later, at 0.71, in the same U.K. survey that was repeated in 2011 in a cross section of 2680 establishments.

[2]Business school professors who teach cases will be acutely aware of the phenomenon, whereby the best student in a discussion group bears a disproportionate share of the workload and is the driving force behind the group's output.

[3]See, for example, Hamm (2010), and also https://betterexplained.com/articles/understanding-the-pareto-principle-the-8020-rule/.

[4]Team production is important in Japan. Using a sample from a 2005 survey of firms (albeit one that excludes the construction and business services industries), Kato and Owan (2011) document in their Table

pan includes the Great Recession, from February 2008 to March of 2009, which affords a source of plausibly exogenous variation in demand for the firm's services (and in particular a reduction in the demand for the firm's working hours). From the perspective of a single firm, the economic crisis is an exogenous event that provides the hours variation necessary to identify the productivity effect of within-team labor reallocation. Focusing on a particular firm and industry holds constant the considerable heterogeneity that would otherwise complicate the interpretation of results in a broader sample. Two factors make this firm an attractive laboratory for studying how team productivity responds to a recession-induced hours reduction. First, the construction industry is strongly sensitive to the business cycle, which makes the recession a particularly effective treatment. Second, Japan is famous for long working hours[5] and for responding to shocks by adjusting hours rather than workers,[6] which implies considerable identifying variation in within-team hours in response to the recession. We explain in section 6 why we expect our results from a single Japanese firm to generalize to the U.S. and other economies.

We find that team productivity increased in this firm following the economic crisis. The question is, why? We focus on two possible mechanisms that, ex ante, could potentially be important. First, team members may become individually more productive following the crisis. This could happen, for example, because shorter work schedules may imbue workers with greater energy and focus per hour, and less exhaustion and

1 that the fraction of responding firms with self-managed teams or cross-functional teams is 0.62. While not specific to Japan, there is also ample evidence that teams are important in the construction industry. For example, the 2011 British WERS data reveal that for the subsample of 103 establishments in the construction industry, the statistic defined in footnote 1 is 0.49. Team production in the construction industry has also attracted considerable attention in the construction management literature (Yap *et al.* 2020). Pressman (2014) describes the increasing importance of teams in the design of construction projects, which is driven by the increasing complexity of projects and by the demands of the marketplace for lower costs and for faster design and construction. In particular, he argues that three recent technologies (i.e., building information modeling, integrated project delivery, and lean construction, which is a strategic methodology borrowed and adapted from the automotive industry) are most effective when they are applied by high-functioning collaborative teams to tackle complex projects.

[5] As noted in Yamamoto (2016), "The length of work hours in Japan stands out among industrialized nations. According to the International Labour Organization (ILO) statistics and other sources, the percentage of workers working long hours (defined as at least 49 hours per week) in recent years is about 10%-16% in North America and Europe, but 22% in Japan."

[6] For institutional reasons, Japan has a longstanding reputation for relying heavily on hours adjustments – as opposed to layoffs and firings – in response to demand shocks, which is exactly what happened during the Great Recession crisis. Japan's Labor Contract Law prohibits "abuse of the right to dismiss", which basically means that a firm cannot lay off its workers unless it has made reasonable efforts to avoid doing so. An implication is that the firm generally cannot lay off its workers when many workers are working long hours, because in that case the court would order the firm to reduce working hours to standard working hours before reducing its employment. The following newsletter contains a concise explanation of how "abuse of the right to dismiss" is defined.

https://www.jurists.co.jp/sites/default/files/newsletter_pdf/newsletter_1701_labor_employment_law.pdf fatigue. Alternatively, fewer projects demanding each worker's attention implies fewer disruptions (Coviello *et al.*, 2014). Second, the labor hours of workers with heterogeneous productivities may be reallocated within the team so that a greater fraction of the team's total hours are contributed by high-productivity workers than before the crisis. The marginal productivity of an additional hour that is assigned to a team of a given size depends on which team member is assigned that hour. As assigned hours increase to meet demand for the firm's output, the time constraints of the team's most productive workers begin to bind, which requires the employer to assign further hours to less productive team members.

Both of these mechanisms are amplified by complementarities in production, which are present in most team settings, including ours. Complementarities arise from interdependence among team members' labor inputs and are one of the main reasons why employers organize production in teams. For example, suppose two team members work closely together and regularly exchange pieces of useful information that enhance each other's productivity. Then an increase in individual worker productivity (for either or both of them) that improves the quality of the information being exchanged will have positive spillover effects on the productivity of the other worker. As another example, suppose that the more productive worker on a two-person team experiences an exogenous increase in their time endowment, perhaps because they just finished some other job that was diverting their attention from the job at hand. As a consequence, this more productive worker can spend more time helping and teaching the less productive worker to avoid bottlenecks. This would result in an increase in team productivity. The size of this increase, again, would depend on the return to helping activities or knowledge spillovers within the team.

After describing the data and production setting in section 2, section 3 presents descriptive empirical evidence of three types. First, consistent with the Pareto Principle of business management, we show that within-team labor allocation tends to be highly concentrated, with a small number of team members contributing the bulk of the hours. A higher within-team concentration of hours is also found to be associated with higher team productivity. Second, we document that the downward adjustment in the labor input during the crisis occurred more for working hours than for employment and that average team productivity increased by nearly 7.5% after the crisis. Third, we present descriptive evidence suggesting that both of the aforementioned mechanisms (i.e., individual productivity and within-team labor reallocation) may contribute to explaining the post-crisis increase in team productivity.

In section 4, we present a theoretical model that explains within-team allocation of labor hours. The model's workers, who differ in their productivities and time endowments, are assigned working hours based on their absolute advantages in production, and they are assigned to tasks based on their capacities (i.e., time endowments). The most productive workers are assigned hours first. When product demand overwhelms those workers' capacities, additional hours are assigned to less productive workers, which decreases average team productivity. We then compute a "pre-crisis" and a "post-crisis" calibration of the model's parameters with data and use it in section 5 to simulate outcomes in both regimes.

Subsequent analyses reveal that the model simulation replicates the empirical patterns observed very well. The calibrated model generates an average team-level productivity increase of 7.6% after the crisis, which is statistically indistinguishable from that found in the real data. The calibration allows us to decompose the total productivity effect into parts due to increased worker-level productivity and within-team labor reallocation. We find that the channel of worker-level productivity increases explains 2.4 percentage points, or 31.6%, of the team-level productivity increase, while labor reallocation explains the remaining 5.2 percentage points, or 68.4%. Therefore, the results suggest that labor reallocation plays an important role in explaining team-level productivity changes. Additionally, the calibrated model successfully generates several patterns that are quantitatively similar to those found in the data, including: (1) the fraction of the team's time accounted for by the team's top hours contributor increases, and team size decreases, after the crisis; (2) the within-team concentration of working hours is positively correlated with team productivity; (3) the joint distribution of output size and the fraction of working hours from the team's top hours contributor is generated from the model-simulated data despite not being explicitly targeted.

Our study contributes by providing new evidence outside of the oft-studied (in the teams literature) manufacturing sector, in particular from the knowledge-intensive, white-collar professional jobs where teamwork is becoming the norm--design, R&D, consulting, accounting, auditing, academic research, etc. Within-team heterogeneity in hours arises in such settings because team members can work in different places, at different times, and for different durations.[7] Economists have been unable to study productivity in such occupations given their idiosyncratic outputs. Our setting and unique data facilitate productivity measurement and analysis because the production process is sufficiently standardized that the total labor required to complete each job is predictable. Moreover, the

[7] In contrast, manufacturing jobs often require workers to be physically and temporally proximate. On an assembly line, for example, complementarities are achieved only when the team members are physically present at the same time, so within-team heterogeneity in working hours (regardless of heterogeneity in abilities) is limited or nonexistent.

value of the output is fixed on each project before teamwork commences. Consequently, productivity depends only on total inputs.

Our focus on within-team heterogeneity in working hours is new. The theoretical model highlights that heterogeneity in hours is a consequence of heterogeneity in team members' individual productivities, where the employer assigns the most productive workers to tasks first, followed by the less productive ones. Although team composition is endogenous in our model, the available talent pool of candidates is randomly drawn, which implies significant variation in the distribution of available skill levels. This team formation process in the model creates a negative correlation between heterogeneity and productivity because less productive teams tend to add more workers from the lower part of skill distribution and more heterogeneous teams in terms of skill level tend to be less productive due to complementarities. Both the actual data and those simulated from the calibrated model exhibit a similar pattern, i.e., team heterogeneity in skills and team size are negatively correlated with productivity.

There is a related literature on team diversity and productivity. Although many dimensions of heterogeneity have been explored, our study investigates the implications for team productivity of worker heterogeneity in productivity and, consequently, in assigned working hours. There are theoretical rationales for both positive and negative team-level productivity effects.[8] Hamilton *et al.* (2012) discuss gains from task coordination and peer learning. Productivity improves when workers' skill levels and task difficulties are optimally matched or when more experienced workers share their knowledge with less experienced ones. On the other hand, when worker heterogeneity makes it difficult to form the team norm or standard, which can be interpreted as the team equilibrium a la Che and Yoo (2001), worker heterogeneity could harm team productivity. When team members are peers who compete with each other for advancement within the organization, additional implications are derived. Classic tournament theory (Lazear and Rosen 1981) predicts that heterogeneity in ability depresses incentives, which would hurt team performance.[9] In contrast, the market-based tournament model of Gürtler and Gürtler

[8] There are also theoretical rationales for both positive and negative team-level productivity effects on dimensions of heterogeneity other than individual productivity (e.g., various demographic characteristics). The positive view is that diversity broadens the set of perspectives and approaches that team members bring to the table, which fosters creativity, scope for complementarities, and ultimately high group performance. The negative view, which is supported by the preponderance of the evidence (Mannix and Neale 2005), is that diversity induces communication challenges and social divisions that hurt group performance. See Lazear (1999) for discussions supporting the positive view and Lang (1986) and Kandel and Lazear (1992) for discussions on the negative view.

[9] The reason is that the high-ability workers need not exert much effort because they are likely to win regardless, and the low-ability workers do not exert much effort because their chances of winning are low regardless.

(2015) shows that the opposite prediction can arise.[10] Empirical evidence favors a positive effect of heterogeneity in ability on team performance. [11]

2 Production setting, data, and measures

The data come from a large Japanese architectural and engineering consultancy firm and include personnel records (from 2011 to 2016) and project management data (from fiscal years 2004 to 2016). The analysis is also informed by in-person interviews that we conducted with seven of the firm's managers and by other less formal communication with the firm's human resource managers.[12] The personnel records cover all employees, including dispatched or contract workers who may be included in the project management data, and include salary and hierarchical ranks that are classified into three levels (manager, senior architect, and junior architect).

Projects consist of multiple phases, called jobs. The job is the unit of observation.[13] Contracts are negotiated separately for each job in a project, with contract terms set before the job begins. Each job is performed in a team of varying size depending on the phase, type of the building, floor space, etc. Teams have an average size of 14 and are assigned to a chief manager who is the person fully responsible for the job and who bears a penalty in the event of quality problems. The chief manager's responsibilities include selecting one team leader, usually a senior architect, to lead daily operations and staffing junior architects who execute tasks (e.g., drawing pictures after the details of the design are confirmed).

The sequence of jobs in a particular architectural project might be as follows: initial planning, schematic design, design development, construction documentation, and supervision of the construction process. In the initial stages, the architects work with the client to discuss the requirements of the building, including the size and the shape. After the basic design is settled, the project team interacts with the client concerning the project's details. For example, the materials for interior finishes must be selected. A large part of the architectural work is construction documentation, where architects produce

[10]The idea is that winning a promotion against a competitive pool characterized by a wide range of talent causes competing employers in the labor market to update their beliefs about the winner's ability to a greater extent than if the worker had prevailed over a level playing field. Workers anticipate large prizes from promotion due to this larger updating, which creates a strong incentive to exert effort to try to win the prize. See also Deutscher *et al.* (2020).

[11]See Hamilton *et al.* (2003, 2012),Franck and Nüesch (2010), Parrotta *et al.* (2014), Chan *et al.* (2014) and Garnero *et al.* (2014).

[12]The seven managers were selected on the basis of the manager effects estimated in Shangguan and Owan (2019), which also contains further details about the data.

[13]Usage of the word "job" here differs from that in either the personnel economics literature or the forthcoming theoretical model. In the context of the data, a job is a phase of a longer-term project.

drawing sets that contain all the details for approval and construction purposes. Finally, the architects work with both clients and contractors at the supervision stage, to ensure that the construction aligns with the design, making design changes when necessary.

The data include two kinds of jobs. External jobs are profit-center jobs that generate revenue. Internal jobs are cost-center jobs that mainly entail administrative responsibilities. Revenue and costs (both labor and non-labor) are observed for each job. Finer components of nonlabor costs are also observed, including material/traveling costs and three types of outsourcing costs. The project management data also include information on the client's identity and industry, type and size of the building being designed, location where the work is conducted, phase of work, contractor selection method, etc.

Information about the entire industry comes from the Current Survey on Orders Received for Construction, conducted by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT). For each ordering industry and each type of construction, the survey reports the annual amount of total orders received by the 50 largest construction companies. The survey has been reported monthly since 1959, and the total construction orders is used as a leading indicator of industrial demand by the Japanese government and think tanks. We aggregate the survey data at the ordering industry level and then link to the data described above as the industry information recorded for each job.

The following definitions of empirical measures index workers and jobs by i and j. Time periods, which are comprised of multiple years, are indexed by t. Specifically, let $t-1$ denote the period before the financial crisis (i.e., years 2005, 2006, and 2007), and let t denote the post-crisis period (i.e., years 2010, 2011, and 2012). We drop the years 2008-2009 from our empirical models due to the difficulty of exactly timing the shock. Let $AfterCrisis_j$ denote a dummy equaling 1 if the starting year for job j occurred in the post-crisis period and 0 otherwise.

2.1 Team productivity

Revenue for job j, Rev_j, is the main output measure. Given that working hours are also observed, a natural measure of team productivity is the ratio of Rev_j and total working hours on job j. However, typically some of the work is outsourced to third parties, whose working hours are unobserved in the data. To render the input and output measures compatible, Rev_j is adjusted using job j's outsourcing costs. Outsourcing costs mostly include the routine tasks that are more efficiently done by a subcontractor and the highly specialized tasks that cannot be completed within the firm (e.g., special inspections that require a specialized license). Let O_j be the outsource ratio for job j, which is calculated as the ratio of outsource costs, $OutsourceCost_j$, and job-level total costs, $Cost_j$. "Adjusted revenue" for job j is then defined as $AdjRev_j = Rev_j \left(1 - O_j\right) = Rev_j - \frac{Rev_j}{Cost_j} OutsourceCost_j$.

The job-level team productivity measure that serves as the main dependent variable is $\ln \frac{AdjRev_j}{h_j}$, or the natural logarithm of the ratio of adjusted revenue to total working hours, h_j, which are defined in the next subsection. Figure 1 shows its distribution for the estimation sample. There is substantial productivity variation across this firm's teams. The standard deviation of $\ln \frac{AdjRev_j}{h_j}$ is about 0.76, implying that a team whose efficiency exceeds the mean by one standard deviation can produce more than 1.5 times the revenue of a team whose efficiency lies one standard deviation below the mean. Figure 1 also plots the density function of a normal random variable with the same mean and standard deviation as in the histogram. The normal distribution approximates the data reasonably well.

The output measure can also be understood by comparing it to the traditional value-added measure. As depreciation is negligible in the current firm, the value added measure for each job can be defined as the sum of job-level profit and labor costs (or, equivalently, Rev_j minus nonlabor costs). As outsourcing costs are the major part of nonlabor costs, adjusted revenue is expected to relate closely to value added. This is verified in Figure 2, where the adjusted revenue for each job is plotted on the horizontal axis, and the value added for each job is plotted on the vertical axis. Compared with the value-added measure, an advantage of the adjusted revenue measure is that it ensures that output is positive. This yields a more intuitive interpretation and ensures that its natural logarithm is defined for every job. As discussed in Syverson (2011), using revenue is the literature's standard approach to measuring output, though it has limitations.[14] We return to this issue at the end of section section 3.2 when addressing the possibility that revenue incorporates price changes that may obscure productivity changes.

2.2 Working hours and other variables

The presumption in this study (as made explicit in the forthcoming model) is that team members' working hours are assigned by the employer, specifically by the chief manager, rather than chosen by the worker.[15] In alternative production settings, the reverse might

[14]One concern is that revenue might reflect changes in product mix. Another is that price variation might reflect differences in market power across producers. In the latter case, revenue may be more reflective of the state of the local output market than of true productive efficiency. See section 2.2. of (Syverson, 2011) for a more detailed discussion.

[15]The chief manager is supervised by the executive committee, but the committee only influences the decision of assigning jobs to chief managers. Chief managers are often assigned to several jobs that they manage simultaneously, and they can decide how much of their attention to devote to each.

be true. Ambiguity concerning which assumption is correct in general is highlighted in Pencavel (2016). This problem of ambiguity is avoided in the present context with single-firm personnel data, in which our interviews with the firm's manager's revealed that hours are assigned to workers by the chief manager.

The managers allocate tasks across workers and plan and monitor how many hours are spent on each task to control labor costs. They are expected to provide advice and support when there is a delay in progress.[16] They also conduct regular internal meetings to communicate about the status of each worker on each project, so as to make better subsequent labor allocation decisions. The data on working hours are available at the worker-job-month level. They are from the project management data and reported by the workers for internal accounting purposes.[17] As referenced in section section 2.1 in the productivity measure's denominator, h_j denotes total working hours on job j. We denote by h_{kj} the total working hours of the worker who ranks k^{th} in terms of total hours within the team. For example, h_{1j} denotes the working hours of job j's rank-1 worker.

Similarly, we use l_{kj} to denote the fraction of hours contributed by the rank-k worker. We also use an alternative measure, l_j^q, which represents the q^{th} percentile of the distribution of hours fractions. For example, for the case of $q = 90$, we define l_j^{90} via the following steps. We first rank team j's members in order (from lowest to highest) by their hours contributions to the team, with the lowest hours attributed to the 0^{th} percentile, and the highest hours attributed to the 100^{th} percentile. Using a simple linear interpolation rule, we then calculate l_j^{90} as a weighted average of the hours fractions of the two workers who straddle the 90^{th} percentile.[18]

[16] The fact that hours are closely monitored by chief managers in this setting eliminates an identification problem that plagues the literature on working hours, i.e., do observed hours reflect workers' preferences or employers' preferences? As discussed in Pencavel (2016) that identification question received attention in the 1960s and 1970s (e.g., Feldstein 1968, Rosen 1969, Abbott and Ashenfelter 1976) but was then largely forgotten for more than four decades as the empirical literature became dominated by labor supply models that implicitly resolved the preceding question in favor of workers' preferences. The present study's operating assumption of employer-determined hours is appropriate in light of our interviews with the firm's managers.

[17] One might wonder whether workers have an incentive to overreport or underreport their hours. Total hours are unlikely to be misreported because excessive hours, which can be electronically double checked, are closely monitored by the management for cost control and health management purposes. Workers, however, might re-allocate their hours among the jobs on which they work. If a job incurs too many costs, workers might re-allocate their hours to another high-margin job, so as to please their bosses. The management, however, strongly discourages such behavior.

[18] Suppose the team has n workers. If the 90^{th} percentile falls between the rank-i and rank-$(i+1)$ workers, i.e., $n_j - (i+1) \leq 0.9(n_j - 1) \leq n_j - i$, the 90^{th} percentile is calculated as the weighted average of the hours shares of the rank-i and rank-$(i+1)$ workers, with the weight for the rank-i worker being $n_j - i - 0.9(n_j - 1) = 0.1n_j + 0.9 - i$. By this definition, 90 percent of the interval $[1, n_j]$ lies to the left of l_j^{90}. This is the standard interpolation rule as shown in method 7 of Hyndman and Fan (1996). When $2 \leq n_j \leq 10$, the 90^{th} percentile is the weighted average of the hours shares of the rank-1 and rank-2 workers.

Let $TeamSize_j$ denote the number of workers engaged in teamwork on job j. It is calculated as the number of unique worker identifiers over the whole production period. In our analysis sample, half the teams have fewer than 10 workers, and fewer than 10 percent have more than 14 workers. The number of team members who continuously work together is about 5, which is calculated as the average number of workers across months. The distribution of team size appears in the top row of Table 1. Let $JobContent_j$ denote a categorical variable (with 22 categories) that we use to control for the type of service in each job j.[19] Let Ind_j denote a vector of 39 dummies indicating the client's industry.[20] Let $AveAgeSt_{1j}$ denote the average age of the rank-1 worker in the starting year of job j.

2.3 Sample selection and summary statistics

Given our use of a revenue-based productivity measure, we focus on external jobs, which are the revenue-generating profit-center jobs. We only include jobs with revenue of at least one million Japanese yen. Since that revenue threshold is rather low across jobs, this restriction essentially excludes failed jobs that do not generate any revenue. Jobs for which the floor area is zero are excluded. These tend to be consulting jobs that differ in nature from design jobs.

The distribution of the outsourcing ratio over all jobs displays two spikes. The first occurs at zero, i.e., many jobs use no outsourcing. The second occurs near one, where the fraction O_j is at or very near one. We restrict the analysis to the sample of jobs with $O_j \leq 0.8$, though using alternative cutoffs yields similar results.

A typical job extends beyond one year, and a large job could last three years. We require jobs to be completed. Although the data cover the period from 2004 to 2016, only jobs that started from 2004 to 2013 are included, to avoid right censoring. The excluded jobs that started before 2004 are expected to be longer jobs. The desirability of dropping observations beyond 2013 is clear from Figure 3, which plots the average duration of jobs by their starting years. A sharp drop is observed in 2014, due to right censoring.

Table 2 reports summary statistics for all variables used in the analysis.

[19] The top 10 categories of $JobContent_j$ cover 96.1% of the number of jobs and 98.4% of revenue in the sample. In decreasing order of revenue, they are: construction documentation (32.6%), design/construction supervision (27.2%), construction supervision (17.0%), design development (13.3%), other (3.3%), schematic design (2.2%), planning & development management (0.9%), other planning (0.8%), construction supervision consulting (0.6%), design/construction supervision consulting (0.5%).

[20] The top 10 client industries cover 65.4% of the number of jobs and 70.6% of the revenue in the sample. They are: Real-estate (21.1%), Education (10.3%), Financial/insurance (9.1%), Transportation (6.2%), Other public interest organizations (5.2%), Municipal government (4.1%), Electronics (4.0%), Others (3.6%), Medical related organizations (3.5%), and Service industry (3.4%).

3 Empirical evidence on teams and productivity

This section documents empirical evidence concerning teams and productivity. Section 3.1 provides evidence on within-team labor allocation and the determinants of team productivity. Consistent with the Pareto Principle of business management, within-team labor allocation is revealed to be heavily concentrated, with a small number of workers accounting for the bulk of the team's hours. Higher concentration is also found to be associated with higher team productivity. Section 3.2 documents the post-crisis increase in team productivity that we seek to explain. Section 3.3 provides evidence suggesting possible channels of influence for the post-crisis productivity increase. Following the crisis, on average, each worker is assigned fewer working hours and participates in fewer jobs, while teams shrink in size and concentrate their total hours more heavily on the team's top workers.

3.1 Within-team labor allocation and the determinants of team productivity

Consistent with the Pareto Principle of business management, our data reveal that within-team working hours are highly concentrated, with a small number of team members contributing the bulk of the hours. Table 1 illustrates the within-team allocation of working hours. The topmost row gives the size distribution of teams (e.g., 4-person teams represent 7.6 percent of the sample). The rows are listed in descending order by the team members' hours contributions, with the highest-ranked worker (i.e., the one who contributes the most hours, who we henceforth refer to as the rank-1 worker) listed first.[21] A striking concentration of within-team labor allocation is revealed. In a 6-person team, the top team member contributes more hours than the 5 others combined. Although the top worker's contribution share of the team's total hours naturally decreases with team size, it remains substantial even in teams as large as 20.[22]

Figure 4 plots the empirical distribution of the fraction of hours contributed by the rank-1 worker. On average, that worker contributes 47 percent of the team's total hours. The distribution exhibits two spikes. One, on the right, corresponds to single-worker teams. The other occurs at about 0.3 and shows that even when excluding single-worker teams, the rank-1 worker often contributes about 30 percent of total hours. Figure 5 plots

[21] For example, in the fourth column, over 64 percent of a 4-person team's hours are contributed by the rank-1 worker, whereas 25 percent are contributed by the rank-2 worker. The rank-3 and rank-4 workers contribute only around 9 percent and 1 percent of total working hours, respectively.

[22] Jobs with team size no greater than 20 comprise 76% of our sample.

the relationship between output size, as measured by $\ln AdjRev_j$, and the fraction of hours contributed by the rank-1 worker. That fraction exhibits substantial variation conditional on output size. The fraction naturally decreases as output size grows, though it remains substantial even for large jobs. Thus, regardless of the size of the job, the rank-1 worker contributes a substantial share of the team's total hours.

We now investigate the extent to which the substantial variation in team productivity that is documented in Figure 1 can be explained by the fraction of hours contributed by the rank-1 worker. Table 4 reports the results from a regression of team productivity on the fraction of the team's hours contributed by the rank-1 worker. The regression shows that the fraction has strong explanatory power; the relationship is significant both statistically and economically. A 10 percentage point increase in the fraction is associated with roughly a 5% increase in team productivity.[23] This relationship suggests the relevance of within-team labor allocation for understanding team productivity.

3.2 Productivity changes surrounding the crisis

Changes in industry demand are shown in Figure 6, which graphs the total orders received annually, in trillions of Japanese yen, as measured by the MLIT survey. Total demand decreased starting in 2008 and did not recover to its pre-crisis level until after 2013. Overall economic conditions, as measured by gross domestic product (GDP), show a similar pattern. According to the World Bank, Japan's GDP in 2010 U.S. dollars decreased from 5.848 trillion in 2007 to 5.471 trillion in 2009 and did not recover to its pre-crisis level until 2013.

The firm experienced a similar drop in demand, as shown in Figure 7. The total revenue from jobs that start in each year suddenly plummeted in 2008 and did not recover to its 2007 level even by 2013. Similarly, Figure 8 shows that the total number of jobs that started in each year decreases from 2007 to 2009 and by 2013 remains lower than its 2007 level.

Figure 9 plots the average adjusted revenue per hour – the natural logarithm of which serves as our dependent variable describing team productivity – for jobs that were started in the year indicated on the horizontal axis and completed by the end of the sample period (i.e., 2016). The plot reveals a trend that decreases until 2008 and then increases until 2012. Retirement and switching to minor roles before retirement of a few highly productive workers from the baby boomer cohort may partly explain the decline in individual

[23] With the estimated coefficient of 0.526, an increase of 0.1 in $\ln l_{1j}$ is associated with an increase in team productivity of about 0.053. The estimated coefficient is even larger when controlling for output size.

productivity.[24] To show evidence for this conjecture, in Figure 10,we plot the average age of rank-1 workers across all jobs that start in each year, normalizing year 2008 to be equal to 1.[25] The average age of the rank-1 worker exhibits a clear decreasing trend that coincides with the pre-crisis trend in team productivity. These patterns suggest the desirability of purging team productivity of age effects, which we do using the following regression model:

$$\ln \frac{AdjRev_j}{h_j} = \beta_0 + \beta_1 AveAgeSt_{1j} + \beta_2 AveAgeSt_{1j}^2 + \phi_j^{Ind} + \phi_j^{JC} + \varepsilon_j. \tag{1}$$

In this regression, ϕ_j^{Ind} denotes client industry fixed effects, and ϕ_j^{JC} denotes job content fixed effects. Controlling for client industry and job type in the preceding regression addresses job heterogeneity. The residuals are regressed on starting year fixed effects. The estimates are plotted in Figure 11, which reveals that the decreasing pre-trend in team productivity disappears, whereas team productivity increases following the crisis.

Table 3 reports the magnitude of the productivity change resulting from the crisis. Column 1 reports the estimate of a simple regression of $\ln \frac{AdjRev_j}{h_j}$ on $AfterCrisis_j$ with no controls. The estimated value of $\Delta \hat{A}_t$ in column 1 is 7.5 percent. This estimate may be biased as a productivity effect of the crisis because the project composition in terms of client industry and job content may change after the crisis. To estimate the conditional productivity change resulting from the crisis, we replace the starting year dummies with the $AfterCrisis_j$ dummy in the preceding regression. Thus, the productivity change resulting from the crisis, conditional on job characteristics, is measured by δ in the following regression:

$$\ln \frac{AdjRev_j}{h_j} = \beta_0 + \delta AfterCrisis_j + \phi_j^{Ind} + \phi_j^{JC} + \varepsilon_j, \tag{2}$$

The estimated δ when the regression includes fixed effects for industry and job content is reported as $\Delta \hat{A}_t$ in column 2 of Table 3, i.e., 6.6 percent. The small and statistically insignificant decrease in the estimated δ between columns 1 and 2 shows that productivity improvement within job categories (rather than a change in the composition of jobs) is

[24] The baby boomers who were born in 1947-49 retired between 2007 and 2009.

[25] More specifically, we estimate a regression of the natural logarithm of the rank-1 worker's age on starting year, industry, and job content fixed effects. We then plot the value of the starting year fixed effect after taking an exponential transformation. The absence of personnel data prior to 2011 means that we do not have valid age data for workers who leave the firm before 2011. Given that the turnover rate is low in this firm and many workers remain there for their careers, we impute the missing ages by assuming that workers who disappear before 2011 do so because they retired at the standard retirement age of 60.

driving the team productivity increase.[26] This estimate increases to 11.5 percent when the quadratic of the average of the rank-1 worker ages is added to the regression, as reported in column 3. This increase in the estimated δ reveals that because the rank-1 workers are on average less experienced after the crisis, to the extent that more experienced rank-1 workers induce higher team productivity, we should attribute a higher productivity enhancing effect to the crisis.

A potential concern with any revenue-based productivity measure is the possibility that revenue incorporates price changes that may obscure productivity changes. This is a well-known and widespread problem in productivity analysis, as discussed in Syverson (2011). In our case, the markup, i.e., the spread between the selling price and the production cost, likely decreased in response to the crisis-induced drop in demand. That decrease may at least partly explain why revenue per hour decreased in 2008, when the shock of the crisis had the largest impact. Following the same logic, the post-crisis increase in revenue per hour may at least partly reflect a recovering markup instead of an improving production technology.

To assess the role of the markup, Figure 12 plots average adjusted revenue per job for jobs that were started in the year indicated on the horizontal axis and completed by the end of the sample. If the change in the markup is the major force that drives productivity, then the average adjusted revenue per job should decrease during the crisis and increase after the crisis. Figure 12 reveals the opposite pattern, suggesting that the productivity increase is not driven by a changing markup.[27] Although this finding that the price is insensitive to a shift in demand might appear surprising, note that projects often last for many years, and this firm had sufficient backlogs of orders to survive the recession. Thus, the firm did not need to cut prices to attract business.

To further investigate the potential role of price changes, we obtained three relevant industrial price indexes from the Bank of Japan and used them to control for price in the

[26] If the pre-crisis and post-crisis definitions are both shortened by a year (i.e., 2006-2007 instead of 2005-2007, and 2010-2011 instead of 2010-2012), column 1 of Table 3 changes to $\Delta\hat{A}_t$ = 10.2 percent with standard error 0.028, column 2 changes to $\Delta\hat{A}_t$ = 8.0 percent with standard error 0.028, and column 3 changes to $\Delta\hat{A}_t$ = 5.1 percent with standard error 0.039. If both periods are lengthened by a year (i.e., 2004-2007 instead of 2005-2007, and 2010-2013 instead of 2010-2012), column 1 changes to ΔA_t = 4.7 percent with standard error 0.021, column 2 changes to $\Delta\hat{A}_t$ = 3.3 percent with standard error 0.020, and column 3 changes to $\Delta\hat{A}_t$ = 11.3 percent with standard error 0.032. The tradeoff is that shorter bandwidths for time periods reduce the sample size, whereas longer ones increase the risk that events unrelated to the crisis may cloud the picture.

[27] A regression analysis that controls for industry and job content dummies shows that while there is an increase of adjusted revenue per job when comparing the pre-crisis period (2005 to 2007) with the crisis period (2008 to 2009), there is no evidence of significant changes between the crisis period and the post-crisis period (2010 to 2012). Given that the regression controls for detailed job characteristics, the difference in the timing of changes further confirms that a changing markup is unlikely to drive the productivity change.

preceding regression model. These indexes are for three separate services: architectural design, civil engineering design, and civil engineering service.[28] The augmented regression yields a considerably larger post-crisis increase in team productivity (i.e., about a 14.6 percent increase in team productivity with a standard error of 0.047) than we document in our main result, which can therefore be considered conservative.

Identification in the preceding regression comes from temporal variation (i.e., the difference before and after the crisis); there is no control group. Given the nearly universal reach of this crisis (particularly in a highly cyclical industry like construction) it is unclear that a valid control group could be defined. Nonetheless, temporal variation alone is still interesting in this case because of the magnitude, nature, and abrupt onset of the crisis. We believe that a single firm's experience in response to this plausibly exogenous major shock is informative, particularly in a highly cyclical industry like construction in which effects of a major downturn are definitely expected. All our evidence points to an abrupt regime change being triggered by the crisis. Moreover, the theoretical model that we calibrate to provide further evidence from simulations adds credence to our results.

3.3 Channels of influence for post-crisis increase in team productivity

We next present evidence suggesting two potential channels of influence for the post-crisis increase in team productivity documented in section 3.2. The first channel increases team productivity through an increase in individual worker productivity, and the second does so via allocating more tasks to more productive workers. Figure 13 reveals that the average monthly working hours (of workers) declined by about 20 hours in 2009, which was likely driven by a sudden drop in demand. The variable does not return to its pre-crisis level even when total revenue recovers in 2013. The decline in working hours is likely to improve the worker's productivity because of less fatigue and more energy. Moreover, Figure 14 plots the average (across workers for each month) number of jobs on which a worker spends a positive amount of time, along with the annual average across all months in each year (which is shown as a solid line). The variable decreases around 2007 to 2010 and never rebounds to its 2006 level. [29] The reduction in the number of

[28] To elaborate, we computed the average of these three indexes, omitting the year 2004 because the civil engineering design index was unavailable for that year. The resulting average price index was included as a control variable in a regression of team productivity that includes $AfterCrisis$ as the independent variable and that incorporates industry and job content fixed effects.

[29] The main reason for the decrease is that the total number of jobs dropped as demand began softening in 2008 when the crisis hit, as cash strapped potential clients delayed pursuing their intended construction projects. According to the national construction statistics, total orders declined by almost 30 percent from the peak in 2008 to the trough in 2010. This overall weakness of the market might have been partially offset by several big urban redevelopment projects that started in mid-2000 for this major architectural design

jobs per worker is supportive of the first channel. Specifically, when the number of jobs declines following the crisis (as documented in Figure 14) the attention of the team's top worker is less likely to be diverted by other jobs, so team productivity should increase. Support for that channel can be found in the work of Coviello *et al.* (2014, 2015) which shows that multi-tasking leads to task juggling that reduces a worker's productivity.

The preceding facts are interesting when combined with the evidence in section 3.1, which reveals the high within-team concentration of hours and that a team's greatest contributor of hours makes a substantial contribution to team productivity. Thus, the second channel of labor reallocation can also potentially contribute to the post-crisis increase in team productivity. Further evidence is suggested in the first two rows of Table 5, which reports estimation results from regressions of the form:

$$Outcome_j = \beta_0 + \beta_1 AfterCrisis_j + \beta_2 \ln AdjRev_j + \phi_j^{Ind} + \phi_j^{JC} + \varepsilon_j, \tag{3}$$

where the first two measures of $Outcome_j$ are $\ln TeamSize_j$ and l_j^{90}, i.e., the 90^{th} sample percentile of the within-team hours fraction, as defined in section 2.2. The first row of Table 5 shows that team size decreases after the crisis. The second row shows that l_j^{90} increases after the crisis. Both patterns are consistent with labor reallocation contributing to the increase in team productivity.

The third row of Table 5 reports estimation results when the dependent variable in the preceding regression is $\ln h_{1j}$, i.e., the natural logarithm of the working hours of team j's rank-1 worker. That worker's hours decreased after the crisis, which suggests a worker-level increase in efficiency (perhaps because the decline in the number of jobs, as shown in Figure 14, allowed workers to better focus their attention with fewer distractions). Unreported regression estimates also reveal decreases in working hours for the team's workers ranked 2 through 5.

Both channels of influence are amplified to the extent that complementarities exist between the hours of different team members. Our discussions with the firm's managers lead us to expect that nontrivial complementarities exist in this setting. For example, some workers need to work out the space design, others need to put on the electricity system, and still others need to design the air conditioning system, etc. Integrating all of these parts can be accomplished more efficiently when the different team members' hours significantly overlap.

In sum, the evidence suggests that post-crisis team productivity increased via two distinct channels: an increase in the productivities of individual team members and within-

company.

team labor reallocation. Both channels are amplified to the extent that complementarities exist, as we believe they do, among the different team members' hours. Although the preceding reduced form models provide evidence consistent with both channels of influence, their relative magnitudes cannot be quantified. In the next two sections, we further analyze the two channels of influence with the aid of a theoretical model that rationalizes all the patterns presented above and permits a quantitative analysis of their relative importance.

4 A theoretical model of labor assignment within teams

The theoretical model has two purposes. The first is to provide an interpretation of the empirical patterns revealed in section 3. The second is to create an analytical framework for quantifying two potential contributors to the increase in team productivity. The model's production process has two stages. In stage 1, teams are formed. In stage 2, working hours are allocated within those teams. We describe these stages in reverse order.

Taking the team's composition and total output as given in stage 2, section 4.1 describes the model's solution for allocating within-team working hours. Section 4.2 extends the model to include stage 1, the team formation process. Section 4.3 explains how the theoretical model helps to interpret the empirical data. Section 4.4 shows how the model permits changes in team productivity to be quantitatively decomposed into the aforementioned two contributors to the team productivity increase, highlighting the role that complementarities play in determining the productivity change. That decomposition, which is conducted in section 5.3, reveals that within-team reallocation of working hours is a quantitatively important driver of changes in team productivity.

4.1 Labor allocation within teams

Consider a single firm (also called the employer) that employs a number of workers (indexed by i) and that operates in a production setting consisting of a set of jobs (indexed by j), each of which is completed by a team of workers. Let ϕ_{ij} and H_{ij} denote worker i's productivity and time endowment, respectively, on job j. Thus, H_{ij} represents the maximum amount of time that worker i could devote to job j. Both ϕ_{ij} and H_{ij} are observed by the employer and assumed to be stochastic draws from a joint distribution. Their correlation in the population is denoted by $\rho_{\phi H}$. For each job, a continuum of tasks, denoted by Ω and indexed by s, must be completed by the team. The total measure of tasks, or $|\Omega|$, is denoted by S. Let Ω_{ij} denote the set of tasks on job j to which worker i is assigned. For simplicity, we assume that each task can be assigned to at most one worker.[30]

Job j's total output is denoted Y_j. Let q_{ijs} denote the amount of output that arises from task s assigned to worker i on job j. A worker with productivity ϕ_{ij} who works for h_{ijs} hours on task s of job j has task-specific output of $q_{ijs} = \phi_{ij} h_{ijs}$. A Cobb-Douglas aggregator over a continuum of tasks combines these task-specific outputs in the following production function for job j:

$$Y_j = \exp\left[\int_{s\in\Omega} \ln\left(q_{ijs}\right)^{\gamma_{js}} ds\right]. \tag{4}$$

The (positive) parameter γ_{js} can be interpreted as the weight that task s receives on job j. We assume for simplicity that tasks are symmetric, i.e., $\gamma_{js} = \frac{1}{S}$.[31]

The employer's problem is to allocate labor within each job by deciding how many worker hours to assign to each task in that job. In this subsection, we take job j's team as given. Job j's team is represented by the set $\left\{\phi_{ij}, H_{ij} | i = 1, ..., n_j\right\}$, with n_j denoting job j's team size. Thus, each worker on job j's team is fully described by their job-specific productivity and time endowment. In section 4.2, we describe the team formation process. Let M_{ij} denote the (endogenous) measure of tasks that the employer assigns to worker i on job j, and let h_{ijs} denote the hours that worker i is assigned by the employer to job j's task s. Because tasks are symmetric, in the optimal solution h_{ijs} must be equal over tasks for the same worker, so the hours spent by worker i on job j, i.e., h_{ij}, satisfy $h_{ij} = M_{ij} h_{ijs}$. Thus, given the value of h_{ijs}, a worker who is assigned to a larger measure of tasks (i.e., M_{ij} is large) will have higher working hours, h_{ij}, on job j.

Each unit of worker i's output on job j, $\phi_{ij} h_{ij}$, is referred to as an "effective labor hour". Let c_{ij} denote the employer's cost per effective hour of worker i's labor on job j. The total wage payment per hour, $w_{ij} \equiv c_{ij}\phi_{ij}$, is assumed to be increasing in productivity. Naturally, this implies that more productive workers earn a higher wage. Moreover, we assume that the cost per effective labor hour is decreasing in productivity, i.e., c_{ij} is lower for workers with higher ϕ_{ij}.[32]

Taking Y_j as given, the employer's problem for job j can be stated as follows:[33]

[30] This is relatively innocuous under the continuous task assumption, because tasks are sufficiently small so that one worker is enough.

[31] The employer's formula for determining how many worker hours to assign to each of the job's tasks can easily be generalized to accommodate asymmetric tasks.

[32] A justification for this assumption is that other firms do not perfectly observe worker's productivity. Therefore, the firm need not pay the full wage to retain the worker.

[33] Equivalently, the employer can maximize production given the budget.

$$\min_{h_{ijs},M_{ij}} \left(\sum_i w_{ij} M_{ij} h_{ijs} \right), \tag{5}$$

subject to

$$Y_j = \prod_i q_{ijs}^{\frac{M_{ij}}{S}},$$

$$q_{ijs} = \phi_{ij} h_{ijs},$$

$$M_{ij} h_{ijs} \leq H_{ij},$$

$$\sum_i M_{ij} = S.$$

That is, the employer assigns hours (for all workers and tasks) to minimize job j's labor costs, subject to both job j's technological constraint and a requirement that each worker's total hours (across all tasks) on job j not exceed the worker's time endowment.

Details of the solution are in Appendix A. The solution for h_{ij} is given by:

$$h_{ijs} = \frac{R_{ij} Y_j}{\phi_{ij}} \tag{6}$$

$$h_{ij} = \frac{M_{ij} R_{ij} Y_j}{\phi_{ij}}, \tag{7}$$

where $R_{ij} \equiv \frac{c_{ij}^{-1}}{\prod_i (c_{ij})^{-\frac{M_{ij}}{S}}}$ is the reciprocal of worker i's unit cost relative to the weighted geometric mean of the reciprocals of unit costs calculated across team members. Intuitively, if the within-team relative marginal cost of allocating effective hour to worker i is higher, then the solution assigns fewer hours to that worker. In general, the mass of tasks assigned to each worker does not have a closed form solution. Total working hours for job $j's$ team are the sum of the hours for each of its members, i.e.,

$$h_j = \sum_i h_{ij} \tag{8}$$

Job j's team productivity, A_j, is

$$A_j \equiv \frac{Y_j}{h_j} = \left(\sum_i \frac{W_{ij}}{\phi_{ij}} \right)^{-1}, \tag{9}$$

where $W_{ij} \equiv M_{ij} R_{ij}$ is the weight attributed to worker i in job j that comes from the labor allocation rule 7. This expression reveals how team productivity is influenced by the two channels discussed in section 3.3, i.e., the individual productivity of team members (as measured by ϕ_{ij}) and the within-team allocation of labor that the employer determines (as measured by W_{ij}).

The intuition underlying the optimal assignment rule is clear from a rearranged Equation (7), i.e., $\sum_i w_{ij} h_{ij} = SY_j \prod_i (c_{ij})^{\frac{M_{ij}}{S}}$. The employer's problem is to find the assignment with the lowest weighted average cost per effective working hour that will achieve a given level of output, Y_j. Given the assumption that more productive workers have a lower cost per effective working hour, the employer starts by assigning the most productive worker (i.e., the one with the highest value of ϕ_{ij}), exhausting her hours (if necessary) before assigning the worker with the second-highest value of ϕ_{ij}, and so on, until the required output, Y_j, is achieved.[34] The assignment of additional workers continues in this fashion until all required tasks on job j are covered by the existing workers, at which point the team size, n_j, is determined.

Optimal within-team labor allocation when team members have heterogeneous productivities requires that individual productivity be at least partially observed by the manager who assigns the hours. The model's assumption that the employer observes ϕ_{ij} is not always reasonable in a team setting. In fact, that is a reason why group-based (as opposed to individual-based) incentive contracts are often used in teams.[35] In our context, however, it is reasonable to assume that the chief manager possesses information about workers' productivities and uses it when assigning hours to workers. This is especially so given that turnover rates at the firm are low for the institutional reasons previously described. Information about workers' productivities is revealed to the employer from the long job tenures and repeated observations of individual workers on a variety of projects.[36]

A potential limitation of the model is that it does not explicitly incorporate the firm's decision to outsource, even though our empirical measure of team productivity is adjusted by the outsourcing cost. The firm may be more likely to outsource when industrial

[34] The implicit assumption is that the remaining employees are assigned to work on profitable activities other than job j.

[35] It is not always the case, however, that individual output is hard to measure in team settings. Moreover, group piece rates are sometimes used in teams even when individual output is easily measured. For example, Koret, the garment manufacturing plant analyzed in Hamilton *et al.* (2003), switched its seamstresses from individual piece-rate pay to a group piece-rate scheme in which they were allowed to self select into teams. At Koret, individual output was easily measured and compensated via individual piece rates prior to the change in the compensation system, which was made for reasons unrelated to the observability of output.

[36] Our interview with the managers confirm the preceding assertions.

demand is high, though two factors mitigate this concern. First, the average outsourcing ratio (i.e., the ratio of total outsourcing costs to the total costs across all jobs that start in a given year) hovers around the same level between the pre-crisis and post-crisis periods (i.e., 2005-2007, and 2010-2012). Second, if the crisis-induced drop in demand leads to less outsourcing, then productivity may increase in those years via complementarities among the team's internal workers, given that they are more often jointly present with less variation in their working hours.[37] Similarly, in the post-crisis period (i.e., 2010-2013), as demand improves, we would expect to see more outsourcing.[38] This should reduce team productivity, via complementarities among internal team members' hours, given that more of the work is being done by outsiders. But this mechanism works against our findings, so to the extent that it is relevant, our empirical result becomes harder to detect in the data.

4.2 Team formation process

We now describe stage 1 of the production process, extending the preceding setup to model team formation. Doing so allows us to generate a sample of jobs, conditional on parameter values, which is necessary for conducting a quantitative decomposition of changes in team productivity.

In stage 1, given the value of Y_j, let $N_j\left(Y_j\right)$ denote the size of the internal candidate pool representing the firm's workers who are available for assignment to job j. We assume that the size of this pool is increasing in the level of required output, Y_j More precisely, letting α_0 and α_1 be strictly positive parameters, we specify the candidate pool size as follows:

$$N_j\left(Y_j\right) = \lceil \alpha_1 Y_j^{\alpha_0} \rceil, \tag{10}$$

where $\lceil x \rceil$ denotes the smallest integer that exceeds x. Note that $n_j \leq N_j$,recalling that n_j denotes the number of workers on job j's team. The parameter α_0 determines the relative number of workers assigned to large and small jobs, whereas α_1 determines the average number of workers assigned to each job. Conditional on Y_j, all randomness comes from the random variables ϕ_{ij} and H_{ij}. Specifically, for each job j we take a set of independent draws from the candidate pool. Given that those draws are independent across jobs, one worker cannot explicitly be assigned to multiple jobs. However, the allocation of a

[37]In fact, in unreported results we find that controlling for start year, industry, and job content fixed effects, there is a negative correlation between the outsourcing ratio and team productivity.

[38]In unreported results, we verified that the outsourcing ratio decreased in 2008-2009 and then increased in 2010-2013.

worker's time across multiple jobs is implicitly captured in H_{ij}, i.e., a particularly low value of this job-specific time endowment can be interpreted as the bulk of worker i's time being consumed by jobs other than j, leaving little left to allocate to job j.

This team formation process renders the analysis tractable while capturing several essential features of reality. In the data, the team formation decision is fully delegated to the responsible manager. Search costs prevent the manager from considering the firm's entire workforce when staffing a job. There is also randomness in workers' availability (e.g., for family-related reasons). The candidate size rule in Equation (10) captures these realities by assigning limited draws to each job and more draws to larger jobs. The randomness in the draws of ϕ_{ij} and H_{ij} reflects managers' constraints. In particular, suppose that more productive workers tend to have lower time endowments (i.e., $\rho_{\phi H} < 0$). This situation captures a trade-off that commonly occurs in practice, i.e., a more productive worker is often demanded by multiple jobs and, thus, has more limited time to spend on each job.

4.3 Connecting the model to data

To see how the model can generate the within-team hours concentration and across-team positive correlation between team productivity and within-team hours concentration, consider a sample of jobs whose output sizes, Y_j, are given. For each job, we use Equation (10) to assign a team of workers in stage 1. We then allocate labor to complete the jobs following the stage-2 analysis in section 4.1. There are two considerations. One is that the optimal assignment rule for allocating labor within the team requires assigning the most productive workers the most work. But that objective is limited by the second consideration, namely workers' time constraints. When $\rho \geq 0$, higher-productivity workers tend to have more liberal time endowments. In that case, given that the optimal assignment rule first exhausts the more productive workers' time, hours concentrate on the most productive workers, and a team with more productive workers has a higher hours concentration, on average.

When $\rho < 0$, the time constraints of the more productive workers tend to be more binding, so that the team's less productive workers are assigned more hours. Even in this case, the model can generate an across-team positive correlation between team productivity and the within-team hours concentration, for the following reason. Consider an increase in ϕ_{ij} for a team member i who is not team j's rank-1 worker and whose time endowment on job j is exhausted. Since $\rho < 0$, the increase in ϕ_{ij} implies that a decrease in H_{ij} is likely, which would have two further implications under the assumption that the expected decrease in H_{ij} is not too large. First, team productivity increases.[39] Second, because $h_{ij} = H_{ij}$, the fractions of hours contributed by team members other than worker i increase. In particular, the hours fraction contributed by the rank-1 worker, i.e., l_{1j}, increases. In other words, the denominator of the hours fraction (i.e., the team's total hours) is decreasing in individual productivity because hours can be assigned more efficiently when individual productivity increases, and the high within-team hours concentration therefore proxies for high average individual productivity within the team. As all workers except (possibly) the least productive one exhaust their time endowment, we expect the preceding channel to be important in driving the across-team correlation between team productivity and the within-team hours concentration.

Which of these cases is most relevant is an empirical question.[40] The forthcoming calibration exercise detailed in section 5 provides a way to empirically estimate which case is more relevant using the information contained in the distribution of the fractions of working hours contributed by each team member. Our data support the case of $\rho < 0$. We show that the simulated data replicate both of the aforementioned features of the empirical data, i.e., the within-team hours concentration and the across-team positive correlation between team productivity and the within-team hours concentration.

4.4 Complementarity and the two channels for team productivity changes

Section 3.3 defined two channels through which team productivity can increase. We now analyze these channels in the context of the theoretical model and show how the presence of complementarities amplifies both channels. Then, in section 5 we calibrate the model parameters and quantify the relative importance of these two channels in explaining the post-crisis increase in team productivity.

To see the role of complementarities in the model, start by observing that (from Equa-

[39]More specifically, from Equation (9), team productivity increases if changes in weights do not offset the increase in ϕ_{ij}. Given that the increase in ϕ_{ij} and the expected decrease in H_{ij} have opposite effects on the total measure of tasks that can be completed by worker i, it is unclear how task assignments would change across workers in the wake of an increase in ϕ_{ij}. However, given that c_{ij} is assumed to be decreasing in ϕ_{ij}, R_{ij} tends to increase in ϕ_{ij}, putting more weight on the worker whose productivity increases.

[40]There may be alternative explanations for the within-team concentration of working hours. For example, suppose that the workers assigned to a job are specialized on certain tasks. Their hours contributions would then be determined by the importance of those tasks within the job. If the required hours contributions are very large for some tasks and very small for others, this could explain the within-teams hours concentration. Although such an alternative may play a role, at this firm the architects are generalists who can perform a wide variety of tasks. Moreover, tasks can often be subdivided into smaller tasks, so that the hours contributions for particular tasks are not necessarily fixed. Finally, unlike the mechanism for hours concentration highlighted by our model, the alternative explanation does not readily explain why within-team hours concentration increased after the crisis.

tion 9) conditional on W_{ij}, the elasticity of team productivity, A_j, with respect to team member i's productivity, ϕ_{ij}, is[41]

$$\frac{\partial \ln A_j}{\partial \ln \phi_{ij}} = A_j \left(\frac{W_{ij}}{\phi_{ij}} \right). \tag{11}$$

Complementarity, which derives from the presence of A_j in this elasticity, amplifies the amount by which team productivity increases in response to a marginal increase in the productivity of an individual team member.[42] This elasticity is positive, i.e., an increase in a team member's productivity increases that team's productivity. Moreover, the individual elasticities for each of the team's workers sum to one, i.e., $\sum_i \frac{\partial \ln A_j}{\partial \ln \phi_{ij}} = 1$. Therefore, the elasticities for the individual workers are all smaller than one. Intuitively, an increase in productivity for an individual will translate to a smaller impact on the team.

Equation 11 implies that in a team with heterogeneous productivities across workers, conditional on labor allocation, a marginal increase in the productivity of the team's least productive worker yields the largest increase in team productivity. As a corollary, if we increase the highest-productivity worker's ϕ_{ij} by a small amount and decrease the lowest-productivity worker's ϕ_{ij} by that same amount, the within-team variance of ϕ_{ij} would increase, and team productivity would decrease.[43] These features of the model capture realistic aspects of team production, i.e., increasing the productivity of less experienced workers reduces the risk of facing bottlenecks, and increasing the within-team dispersion of individual productivities makes communication/coordination harder.

Finally, a team is more productive if more productive workers are assigned more tasks. Given that the measure of tasks assigned, or M_{ij}, is endogenous, labor reallocation in-

[41] Any change in the weight after a change in individual productivity either comes from reallocation of tasks (i.e., a change in M_{ij}) or an adjustment of hours due to a change in the marginal cost per effective hour (i.e., a change in R_{ij}).Thus, the current exercise can be viewed as the partial effect of an individual productivity change, holding the indirect effects constant.

[42] More precisely, let i_1 and i_2 denote two workers on the team for job j. Conditional on M_{ij}, the cross partial derivative of team productivity with respect to those two workers' individual productivities is

$$\frac{\partial^2 \ln A_j}{\partial \ln \phi_{i_2 j} \partial \ln \phi_{i_1 j}} = \frac{M_{i_1 j} R_{i_1 j} M_{i_2 j} R_{i_2 j}}{\phi_{i_1 j} \phi_{i_2 j}} A_j^2 > 0.$$

Despite this positive cross partial derivative, the second-order elasticity with respect to worker $i's$ own productivity is negative, i.e., $\frac{\partial^2 \ln A_j}{\partial \ln \phi_{ij}^2} < 0$. In fact, it is easily verified that the second-order elasticities sum to zero, i.e., $\sum_{i_2} \frac{\partial^2 \ln A_j}{\partial \ln \phi_{i_2 j} \partial \ln \phi_{i_1 j}} = 0$.

[43] In the reverse case, i.e., we decrease the highest-productivity worker's ϕ_{ij} by a small amount and increase the lowest-productivity worker's ϕ_{ij} by that same amount, the within-team variance of ϕ_{ij} would decrease, and team productivity would increase.

duces a larger increase in team productivity than would occur if individual worker productivity were to increase while maintaining the original labor allocation. For example, if each worker's productivity were to improve uniformly by a certain proportion, then the increase in team productivity would exceed this proportion because of task reallocation. The labor reallocation effect also depends on the second-order derivatives. To illustrate, note that the elasticity with respect to the weight W_{ij}, is

$$\frac{\partial \ln A_j}{\partial W_{ij}} = -\frac{1}{\phi_{ij}} A_j. \tag{12}$$

Let ΔW denote a discrete change that reallocates the weights, reducing the task assignment of a less productive worker, i_1, by some amount and shifting it to a more productive worker, i_2. The resulting change in logarithmic team productivity is $\Delta \ln A_j = \Delta W A_j \left(\frac{1}{\phi_{i_1 j}} - \frac{1}{\phi_{i_2 j}}\right)$, which is positive given that $\phi_{i_2} > \phi_{i_1}$. Moreover, this productivity increase is larger when other team members are more productive (i.e., when A_j is higher), which reflects complementarity.

5 Empirical analysis of theoretical model

We next conduct empirical analysis motivated by the theoretical model. Section 5.1 calibrates the model's parameters, section 5.2 examines the model fit and provides evidence that supports the model's mechanisms, and section 5.3 quantifies the relative contributions of the two channels of influence on team productivity changes.

5.1 Calibration of the model's parameters

The empirical evidence from section 3.1 reveals that the crisis induced a structural change in team productivity in the workplace. To capture this change quantitatively using the theoretical model, and to decompose it into parts due to increased worker-level productivity and within-team labor reallocation, at least some of the model's parameter values must change over time. We now summarize how we assign numerical values to the model's parameters. For brevity, our discussion highlights the main procedures and results, deferring the details to Appendix B. We begin by imposing distributional assumptions. For each job j, the draws of candidates are assumed to be independent and identically distributed. Specifically, candidate i's productivity parameter, ϕ_{ij}, and time endowment parameter, H_{ij}, are assumed to be jointly log-normally distributed, i.e.,

$$\begin{pmatrix} \ln \phi_{ij} \\ \ln H_{ij} \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right), \tag{13}$$

where $\boldsymbol{\mu} = \begin{pmatrix} \mu_\phi \\ \mu_H \end{pmatrix}$ is a vector of means, and the covariance matrix Σ contains two variance parameters (σ_ϕ and σ_H) and a correlation parameter, $\rho_{\phi H}$.

We define a pre-crisis sample of jobs (i.e., those starting from 2005 to 2007) and a post-crisis sample of jobs (i.e., those starting from 2010 to 2012). For both time periods, we must assign values to the following seven parameters:

$$\left(\mu_\phi, \sigma_\phi^2, \mu_H, \sigma_H^2, \rho_{\phi H}, \alpha_0, \alpha_1\right).$$

We calibrate three of these parameters ($\mu_H, \sigma_H^2, \alpha_1$) directly from the data in the two samples, yielding pre-crisis and post-crisis values for each parameter. Parameter values for $\left(\mu_H, \sigma_H^2\right)$ are assigned using the empirical average and standard deviation of the hours of the workers, separately before and after the crisis. Motivated by the fact that the largest observed team size is close to 100, the parameter α_1 is chosen such that the biggest job gets assigned 300 draws, both before and after the crisis. Given those parameter values, we next generate simulated data sets (both pre and post-crisis) that we use to assign values for ($\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0$) using the method of simulated moments. That is, in both the pre and post-crisis samples, values of ($\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0$) are chosen to minimize a distance function that is the sum of the squared deviations between moments computed using the empirical data and the corresponding moments computed using the simulated data. Our target moments are calculated from the distributions of team productivity, output size, and within-team labor allocation.

The preceding steps deliver pre and post-crisis values for all seven parameters, as shown in Table 6. The calibrated parameters reveal that the crisis induces an increase in μ_ϕ and a decrease in σ_ϕ. The mean of the time endowment, μ_H, slightly increases after the crisis, and the standard deviation of the time endowment, σ_H, does not change significantly. The calibrated $\rho_{\phi H}$ increases after the crisis. Its negative sign is consistent with the intuition that more productive workers tend to be time constrained.[44] Together with the fact that μ_H increases after the crisis, a higher value of $\rho_{\phi H}$ implies that more

[44] This can also be understood by observing that, from the standpoint of a social planner who is optimally allocating labor, the implicit price of productive workers will be higher because of their higher marginal output.

hours are allocated to more productive workers after the crisis. The value of α_0 is almost the same before and after the crisis. Figure 15, which plots the implied density function of $\ln \phi_{ij}$ before and after the crisis, illustrates the productivity increase arising from the crisis. The increase in worker-level productivity happens at the lower and middle parts of the distribution, whereas the change is smaller for high-productivity workers.

The motivation for our calibration procedure can be understood as follows. In the model, team productivity relates closely to a weighted sum of the team members' individual productivities, as described in Equation (9). Thus, the observed team productivity distribution can be used to infer the worker productivity distribution. By focusing on the entire productivity distribution, we avoid the difficulty of using team output to infer team members' individual contributions to that output. Moreover, the observed hours contributed by each worker naturally help us to determine the parameters of the time endowment distribution because the model's optimal assignment rule predicts that all workers (except for the least productive one) exhaust their time endowments.

The within-team hours distribution helps us to identify the correlation between productivity and the time endowment because, conditional on other parameter values, labor concentration is an increasing function of $\rho_{\phi H}$. For example, a job's hours are more concentrated when highly-productive workers have more time to devote to the job. In a sample with heterogeneous output sizes the relative productivity between larger and smaller jobs helps to determine the value of α_0. This is because the parameter α_0 determines how many candidates are allocated to large jobs relative to small ones. If the number of talented workers is proportionally increasing, for example, then output size does not substantially affect team productivity. Finally, the value of α_1 does not have real effects on the model other than setting an upper bound for team size because other parameter values are chosen accordingly to match the empirical moments. Therefore, we choose α_1 such that the number of draws for the job with the highest output size is somewhat higher than the observed team size.

5.2 Model fit and validation

Concerning model fit, Figures 16 to 19 show the simulated and empirical density functions of each targeted distribution. In all figures, the dashed lines correspond to the empirical data and the solid lines to the simulated data. Figures 16 and 17 show that the simulated data fit the distribution of team productivity and output size well for both the before-crisis and the after-crisis sample. The distributions of the hours fractions, shown starting from Figure 18 to Figure 19, also exhibit good fit. Figures 20 and 21 show the fit of the cumulative hours contributed by the five workers who contribute the most hours. Even though the distributions of working hours are not explicitly targeted, the model explains them reasonably well. Finally, Figures 22 and 23 show the scatter plots of the rank-1 worker's hours fraction, conditional on output size, both before and after the crisis. Even though the conditional distribution is not explicitly targeted, the simulated data closely approximate the shape of the plots.

Concerning model validation, our goal is to show that the model replicates the empirical patterns that we document in section 3 and to demonstrate that our claims in section 4.3 are corroborated by the simulation. We first estimate the correlation between team productivity and the natural logarithm of the rank-1 hours fraction using the simulated data. The results are reported in Table 7. Similar to Table 4, we find a significant positive correlation between the rank-1 hours fraction and team productivity. Moreover, taking advantage of the simulated data, we verify that the teams with a higher rank-1 hours fraction tend to have higher average productivity and, therefore, higher team productivity. We then estimate regressions of the following form to examine job-level changes in several dependent variables in response to the crisis:

$$Outcome_j = \beta_0 + \beta_1 AfterCrisis_j + \beta_2 \ln Y_j + u_j, \tag{14}$$

using four measures of $Outcome_j$: (1) $\ln A_j$, the natural logarithm of productivity;[45] (2) $\ln TeamSize_j$; (3) $\ln l_j^{90}$; and (4) $\ln h_{1j}$. The latter three measures appeared earlier in Table 5, based on the empirical data. Measures 2 and 3, which are defined and used in Table 5, are for the purpose of validating the model mechanisms that affect team productivity. When measure 4 is the outcome variable, as in section 3 based on the empirical data, we use the jobs with team sizes no smaller than five.

Table 8 reports estimates of post-crisis changes in the job-level variables. The first column shows that in the simulated data, job-level productivity increases by 7.5%. This is the same as in the first column in Table 3 because we included the variable as one of the targets. The effect of labor reallocation is reflected in the second and third columns of Table 8, where it is shown that $\ln l_j^{90}$ increases, and $\ln TeamSize_j$ decreases, after the crisis, as is true in the empirical data (Table 5). These results are consistent with the firm relying on smaller teams in the wake of a crisis-induced reduction in demand. Overall, the calibrated model successfully reproduces the qualitative and quantitative patterns in the data, thereby providing support for the relevance of the model's mechanisms described in section 4.4.

[45]Output size is not controlled in this regression.

5.3 Quantitative decomposition of changes in team productivity

We now quantify the relative importance of the two channels that affect team productivity described in section 4.4, using an approach directly based on the model.[46] We start by calibrating the model and generating counterfactual jobs and teams in the post-crisis environment for each job in the pre-crisis sample. By performing a job-to-job comparison, we alleviate the concerns posed by confounding job-specific factors, such as whether differences in output size may drive the team productivity difference. For each job in the pre-crisis sample, we generate ten counterfactual jobs. In each of those counterfactual jobs, a team is formed and production occurs in the post-crisis environment.[47] The counterfactual jobs are indexed by p, so A_{jp} denotes team productivity for counterfactual simulation p of job j.

A complication that arises when using the model directly is that labor allocations that are feasible in the actual job may be infeasible in the counterfactual job due to time constraints.[48] To avoid this, we must find a counterfactual job that is as "close" as possible to the original job but that maintains a feasible labor allocation that violates no time constraints. We have designed an algorithm to accomplish that. After applying the algorithm to identify such a counterfactual job, we use the theoretical model to decompose the team productivity change into the two channels previous discussed.

To find the nearest feasible allocation for each counterfactual job, we solve the follow-

[46]As observed by a referee, this decomposition of the increase in team productivity evokes the well-known decomposition of the increase in individual worker productivity at Safelite AutoGlass (Lazear (2000)), though with some important differences. Specifically, Lazear's decomposition results from many workers' decisions to select across firms in response to an individual employer's decision to change its compensation system to induce incentives, whereas our decomposition results from an individual employer's assignment of its workers' hours across tasks within teams in response to an exogenous change in demand for the firm's services.

[47]More precisely, output size is held constant, and the job duration is predicted from quantile regressions that use the natural logarithm of the job duration as the dependent variable, with $AfterCrisis_j$ and the natural logarithm of output on the right-hand side, i.e., $\ln T_j = \beta_0 + \beta_1 AfterCrisis_j + \beta_2 \ln Y_j + u_j$. The ten chosen quantiles are equally spaced, starting at the 10^{th} percentile and ending at the 90^{th} percentile. Teams are formed following the joint distribution of ϕ_{ij} and H_{ij} that we calibrated using the after-crisis sample. We find quantitatively similar results when predicting job duration using a linear regression.

[48]Consider a simple illustration in a two-worker team. Holding the less productive worker's productivity constant, consider a 20% productivity increase for the more productive worker that occurs simultaneously with her time constraint becoming more binding. The resulting team productivity increase may be less than 10% (which is the increase in the average of the two individual productivities, i.e., 0% and 20%). This would lead to a conclusion that the effect of individual productivity exceeds 100%, while the labor reallocation effect is negative.

ing problem by choosing the task allocation profile, $\overline{M}_{ijp}$:

$$\min_{\overline{M}_{kjp}} \sum_{k=1}^{\min\left(n_j, n_{jp}\right)} \left(\overline{M}_{kjp} - M_{kj}\right)^2, \tag{15}$$

subject to

$$h_{ijp} = \overline{M}_{ijp} h_{ijps} \leq H_{ijp},\ for\ i \in \left\{1, 2, ..., n_{jp}\right\}, \tag{16}$$

$$\sum_k \overline{M}_{kjp} = S, \tag{17}$$

where h_{ijps} is defined as in Equation (6). Details of the algorithm used to solve the preceding problem are in Appendix C. The team sizes might differ between the actual and counterfactual teams. In such cases, we can only match the measure of tasks for the set of workers with the same rank indexed by k, as shown in the objective function. To minimize error, we sort the workers in team j by M_{kj} from the highest to the lowest, and sort the workers in the counterfactual team jp by the potential of task completion from the highest to the lowest.[49] This sorting method ensures that the errors generated from the matching of high contributors are smaller, which results in lower overall errors.[50]This gives us the counterfactual team productivity with the nearest feasible labor allocation $\overline{A}_{jp}$.

We then calculate the following decomposition of the average team productivity change

$$\frac{1}{JP} \sum_{j,p} \left(\ln A_{jp} - \ln A_j\right) = \frac{1}{JP} \sum_{j,p} \left(\ln A_{jp} - \ln \overline{A}_{jp}\right) + \frac{1}{JP} \sum_{j,p} \left(\ln \overline{A}_{jp} - \ln A_j\right), \tag{18}$$

where the first term on the right-hand side accounts for the effect of labor reallocation and the second term for the individual productivity change.

Let $\ln \overline{A}_{jp}$ denote the natural logarithm of team productivity in the counterfactual team that results from our algorithm (i.e., nearest feasible labor allocation). Table 9 reports the decomposition results. The effect of labor reallocation is calculated as 0.052, or 68.4% (i.e., 0.052 / 0.076) of the average team productivity change, while the average individual productivity change is 0.024, or 31.6% of the average team productivity change.

[49]From Equation 21 in Appendix B, this implies sorting workers by $c_{ij}\phi_{ij}H_{ij}$ from the highest to the lowest.

[50]Searching over all permutations and finding the global minimum is computationally very expensive. Alternative sorting methods (e.g., by hours, by productivity, and randomly) yield higher errors than using our chosen method.

6 Conclusion

This study opened the black box of white-collar team productivity and revealed what drives it in teams of knowledge workers in a representative firm. One of our key contributions is a quantitative decomposition of the crisis-induced increase in team productivity into two channels of influence. The decomposition is based on a new theoretical model that describes the within-team allocation of labor hours and its implications for team productivity. The model might be enriched in a number of interesting directions in future work, e.g., incorporating features like problem solving, coordination, and peer learning, all of which are relevant in team settings. The Cobb-Douglas technology, though it performed well and was able to closely match some key features of our data, might also be generalized in future work, which would allow additional functions of teams to be fruitfully addressed that we abstract from in this analysis.

Our results are summarized as follows. The financial crisis led team productivity to increase by nearly 7.5%, arising from both an increase in individual labor productivity and a within-team reallocation of labor. The theoretical model successfully replicates the distributions of team productivity and labor allocation. A decomposition based on the theoretical model reveals that 2.5 percentage points come from an increase in individual productivity, and the remaining 5 percentage points of the team productivity increase come from labor reallocation. Additionally, we find evidence that within-team working hours are heavily concentrated, with a large fraction of the work being done by a small number of workers, particularly the one worker who invests the most hours. The fraction of time spent by the team's member with the greatest hours contribution is found to be positively associated with team productivity.

In a study that is based on a single firm within a country that has distinctive labor market institutions, it is natural to question the extent to which the analysis and results might generalize. Potential threats to external validity arise for several reasons. This firm might not be representative of architectural and engineering consultancy firms (even within Japan), the industry itself may be idiosyncratic even if this firm is representative of the industry, the institutional environment is specific to Japan, the global financial crisis might be an idiosyncratic example of a major recession, etc. While such issues are acknowledged, a number of factors mitigate them and lead us to expect our results to be relevant for teams of knowledge workers within the U.S. and other economies.

It is true that Japan and the U.S. respond differently to negative demand shocks. Specifically, downward adjustments in employment (as opposed to hours) are relatively more common in the U.S. than in Japan.[51] This institutional difference does not pose a major threat to external validity, because our primary objective is to study not productivity over the business cycle but rather the productivity effects of within-team allocation of labor hours, using the Great Recession as a convenient and plausibly exogenous treatment that induces significant temporal variation in hours. For that purpose, Japan's distinctive emphasis on hours adjustments is more of a plus than a liability.[52] Concerns that the Great Recession might be idiosyncratic among recessions are similarly allayed, because our main interest is not in the productivity effects of recessions per se.

The highly educated workers in our sample are comparable to salaried (as opposed to hourly) workers in the U.S., for whom layoffs would not be used but hours would fall in recessions. Our empirical evidence that highly productive teams of such educated knowledge workers have a high concentration of working hours also supports the oft-cited Pareto Principle of business management, which is mainly discussed anecdotally with reference to employment settings outside of Japan (particularly the U.S.).[53] And again, as discussed in the introduction, team production is a widespread phenomenon internationally as well as in Japan and is of increasing importance in the construction industry. Thus, the employer's problem of within-team labor allocation is relevant to other firms within and outside of this industry.

After multiple extensive conversations with the firm's management, nothing about our firm stands out to us as being particularly unusual or idiosyncratic in its characteristics, services provided, industry, business strategy, management practices, production process, position within the product and labor markets, use of teams, etc., that would lead us to worry that the white-collar team productivity issues that we study here are peculiar to this firm. Thus, while future research in other production settings is desirable, we

[51] While U.S. employers rely on layoffs and firings (particularly in manufacturing), they rely even more on natural attrition. That is, the relatively high quit rates in the U.S., even during recessions, allow firms to achieve timely downward adjustments in their labor inputs simply by not replacing departed workers (Lazear and Spletzer, 2012). Such a labor supply response (via quits) is unimportant in Japan, where salaried workers will not be deciding to exit the firm during a recession.

[52] Setting aside this point, Japan's institutional features that differentiate it from the U.S. and other industrialized economies should not be overstated. Even in Japan, adjustments in (non-standard contract) workers occur. Japan's prohibition of the "abuse of the right to dismiss" applies only to regular workers. Terminating contracts with workers hired under fixed-term contracts is not prohibited. Moreover, even in lightly regulated labor markets in which it is easier than in Japan to shed workers, downward adjustments in hours occur and are typically among the first employer responses in a recession.

[53] Highly right-skewed pay distributions, within and even across firms, might be interpreted as indirect evidence of the Pareto Principle of business management. For example, technology firms in the U.S. (like Google) are known for having highly right-skewed pay distributions in which some workers get extremely high salaries and bonuses to reward their high performance. Shaw (2009) provides such evidence for software workers. Such high performance levels from the stars plausibly require very long hours, implying a highly right-skewed hours distribution.

anticipate that such inquiry should be corroborative. In particular, we expect our results from Japan to generalize to the U.S. and other industrialized economies, and specifically to large segments of the white-collar labor force, such as knowledge workers employed in R&D, consulting, law, accounting, and finance.

References

ABBOTT, M. and ASHENFELTER, O. (1976). Labour supply, commodity demand and the allocation of time. *The Review of Economic Studies*, **43** (3), 389–411.

CHAN, T. Y., LI, J. and PIERCE, L. (2014). Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science*, **33** (4), 463–484.

CHE, Y.-K. and YOO, S.-W. (2001). Optimal incentives for teams. *The American Economic Review*, **91** (3), 525–541.

COVIELLO, D., ICHINO, A. and PERSICO, N. (2014). Time allocation and task juggling. *American Economic Review*, **104** (2), 609–623.

—, — and — (2015). The inefficiency of worker time use. *Journal of the European Economic Association*, **13** (5), 906–947.

DEUTSCHER, C., GÜRTLER, M., GÜRTLER, O. and DEVARO, J. (2020). Firm choice and career success - theory and evidence. *European Economic Review*, **127**, 103470.

DEVARO, J. (2006). Teams, autonomy, and the financial performance of firms. *Industrial Relations: A Journal of Economy and Society*, **45** (2), 217–269.

FELDSTEIN, M. S. (1968). Estimating the supply curve of working hours. *Oxford Economic Papers*, **20** (1), 74–80.

FRANCK, E. and NÜESCH, S. (2010). The effect of talent disparity on team productivity in soccer. *Journal of Economic Psychology*, **31** (2), 218 – 229, the Economics and Psychology of Football.

GARNERO, A., KAMPELMANN, S. and RYCX, F. C. (2014). The heterogeneous effects of workforce diversity on productivity, wages, and profits. *Industrial Relations: A Journal of Economy and Society*, **53** (3), 430–477.

GÜRTLER, M. and GÜRTLER, O. (2015). The optimality of heterogeneous tournaments. *Journal of Labor Economics*, **33** (4), 1007–1042.

HAMILTON, B. H., NICKERSON, J. A. and OWAN, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, **111** (3), 465–497.

—, — and — (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of Participatory and Labor-Managed Firms*, vol. 13, pp. 99–138.

HAMM, T. (2010). *Applying the 80/20 Rule to Your Employees*. Americanexpress, https://www.americanexpress.com/en-us/business/trends-and-insights/articles/applying-the-8020-rule-to-your-employees-1/.

HYNDMAN, R. J. and FAN, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, **50** (4), 361–365.

KANDEL, E. and LAZEAR, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, **100** (4), 801–817.

KATO, T. and OWAN, H. (2011). Market characteristics, intra-firm coordination, and the choice of human resource management systems: Theory and evidence. *Journal of Economic Behavior & Organization*, **80** (3), 375–396.

LANG, K. (1986). A language theory of discrimination. *The Quarterly Journal of Economics*, **101** (2), 363–382.

LAZEAR, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal*, **109** (454), C15–C40.

— (2000). Performance pay and productivity. *American Economic Review*, **90** (5), 1346–1361.

— and ROSEN, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, **89** (5), 841–864.

— and SPLETZER, J. R. (2012). Hiring, churn, and the business cycle. *American Economic Review*, **102** (3), 575–579.

MANNIX, E. and NEALE, M. A. (2005). What differences make a difference?: The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, **6** (2), 31–55, pMID: 26158478.

PARROTTA, P., POZZOLI, D. and PYTLIKOVÁ, M. (2014). Labor diversity and firm productivity. *European Economic Review*, **66**, 144 – 179.

PENCAVEL, J. H. (2016). Whose preferences are revealed in hours of work? *Economic Inquiry*, **54** (1), 9–24.

PRESSMAN, A. (2014). *Designing Relationships: The Art of Collaboration in Architecture*. Routledge.

ROSEN, S. (1969). On the interindustry wage and hours structure. *Journal of Political Economy*, **77** (2), 249–273.

SHANGGUAN, R. and OWAN, H. (2019). How good managers steer their projects: Using value-added measures of manager quality.

SHAW, K. (2009). Insider econometrics: A roadmap with stops along the way. *Labour Economics*, **16** (6), 607–617.

SYVERSON, C. (2011). What Determines Productivity? *Journal of Economic Literature*, **49** (2), 326–365.

WALES, D. J. and DOYE, J. P. (1997). Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *Journal of Physical Chemistry A*, **101** (28), 5111–5116.

YAMAMOTO, I. (2016). Can Japan's long work hours be corrected? adopting regulations to cap total work hours. *Nihon Keizai Shimbun*.

YAP, J., LEONG, W. and SKITMORE, M. (2020). Capitalising teamwork for enhancing project delivery and management in construction: empirical study in malaysia. *Engineering, Construction and Architectural Management*, **ahead-of-print**.

	% of sample	0.7	7.5	8.4	7.6	6.0	5.4	5.0	4.4	4.6	3.6	3.0	2.8	3.0	2.3	1.9	2.1	2.1	1.8	1.6	1.7
	Team Size	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank	1	1.00	0.94	0.75	0.64	0.57	0.53	0.49	0.45	0.43	0.42	0.40	0.38	0.39	0.36	0.37	0.35	0.34	0.31	0.33	0.31
	2		0.06	0.23	0.25	0.25	0.24	0.24	0.23	0.23	0.22	0.22	0.22	0.20	0.20	0.20	0.20	0.19	0.20	0.18	0.18
	3			0.02	0.09	0.12	0.13	0.13	0.14	0.14	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.12	0.13
	4				0.01	0.05	0.06	0.07	0.08	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	5					0.01	0.03	0.04	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.07	0.07	0.07
	6						0.00	0.02	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.05	0.05
	7							0.00	0.01	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04
	8								0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03
	9									0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.03
	10										0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02
	11											0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	12												0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	13													0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
	14														0.00	0.00	0.00	0.00	0.00	0.01	0.01
	15															0.00	0.00	0.00	0.00	0.00	0.00
	16																0.00	0.00	0.00	0.00	0.00
	17																	0.00	0.00	0.00	0.00
	18																		0.00	0.00	0.00
	19																			0.00	0.00
	20																				0.00

Note:Average share of hours contributed by each worker on the team, conditional on the rank of total hours contributed and the team size. The numbers at the top of the table are the shares of each team size in the sample (e.g., 6-person teams represent 5.4 percent of the sample).

Table 1: Allocation of working hours across team members

	count	mean	standard deviation	min	25%	50%	75%	max
$\frac{Rev_j}{h_j}$	6492	27,222.813	148,789.105	499.645	9,513.217	14,324.449	23,631.684	6,450,000.000
$\frac{AdjRev_j}{h_j}$	6492	21,324.002	148,326.713	352.558	7,147.623	10,190.722	15,973.354	6,450,000.000
Rev_j	6492	40,767,358.356	87,589,608.484	1,000,000.000	3,200,000.000	10,000,000.000	37,500,000.000	1,148,320,000.000
$AdjRev_j$	6492	28,538,545.247	58,592,021.352	252,449.133	2,452,058.221	7,403,649.564	27,410,674.736	778,876,558.156
h_j	6492	3,215.822	6,816.450	1.000	185.875	729.750	3,137.125	106,801.500
$TeamSize_j$	6492	14.803	13.897	1.000	5.000	10.000	20.000	98.000
O_j	6492	0.236	0.219	0.000	0.032	0.188	0.378	0.800
h_{1j}	6492	794.095	1,197.948	1.000	107.500	310.000	992.625	12,954.500
l_{1j}	6492	0.474	0.246	0.051	0.279	0.411	0.625	1.000
l_j^{90}	6492	0.340	0.266	0.019	0.125	0.260	0.485	1.000
$AveAgeSt_{1j}$	6492	43.449	1.334	41.575	42.189	43.047	44.411	45.813

Note: Summary statistics for all variables in the analysis, as defined in section 2.

Table 2: Summary statistics

Pre-crisis definition	Post-crisis definition	(1) $\Delta\hat{A}_t$	(2) $\Delta\hat{A}_t$	(3) $\Delta\hat{A}_t$
$2005 \leq StartYear \leq 2007$	$2010 \leq StartYear \leq 2012$	0.075^{***}	0.066^{***}	0.115^{***}
		(0.024)	(0.023)	(0.033)
Industry, Job Content fixed effects		No	Yes	Yes
Time trend of rank-1 worker's age		No	No	Yes
Sample size		4011	4011	4011
Adj. R^2		0.002	0.099	0.100

Note: $\Delta\hat{A}_t$ is the productivity change after controlling for the fixed effects indicated in the table. Standard errors are reported in parentheses. Statistical significance at the 1% level on a two-tailed test is indicated by ***.

Table 3: Change of productivity

	$\frac{AdjRev_j}{h_j}$
$\ln l_{1j}$	0.526^{***}
	(0.024)
Sample size	4011
Adj. R^2	0.197

Note: Industry and job content fixed effects are controlled. Standard errors are reported in parentheses. Statistical significance at the 1% level on a two-tailed test is indicated by ***.

Table 4: Explaining team productivity using hour fraction

Outcome	Change after crisis	Sample size	Adj. R^2
$\ln TeamSize_j$	-0.043^{***}	4011	0.738
	(0.016)		
$\ln l_j^{90}$	0.052^{***}	4011	0.693
	(0.016)		
$\ln h_{1j}$	-0.090^{***}	3060	0.801
	(0.022)		

Note: Estimation results of Equation (3). Sample includes jobs with $2005 \leq StartYear \leq 2007$ or $2010 \leq StartYear \leq 2012$. Regressions control for industry and job content fixed effects. Standard errors are reported in parentheses. Statistical significance at the 1% level on a two-tailed test is indicated by ***.

Table 5: Change of job-level variables after crisis

Sample	α_0	α_1	μ_ϕ	σ_ϕ	μ_H	σ_H	$\rho_{\phi H}$	error
Pre-crisis: $2005 \leq StartYear \leq 2007$	0.519	0.008	7.942	1.205	-0.732	1.432	-0.600	0.016
Post-crisis: $2010 \leq StartYear \leq 2012$	0.521	0.008	8.265	0.930	-0.681	1.450	-0.524	0.013

Note: Calibrated parameter values. The error is the minimized value of the objective function.

Table 6: Calibrated parameters

	A_j
$\ln l_{1j}$	0.442^{***}
	(0.017)
Sample size	3689
Adj. R^2	0.155

Table 7: Validating the correlation between team productivity and the rank-1 worker hours fraction

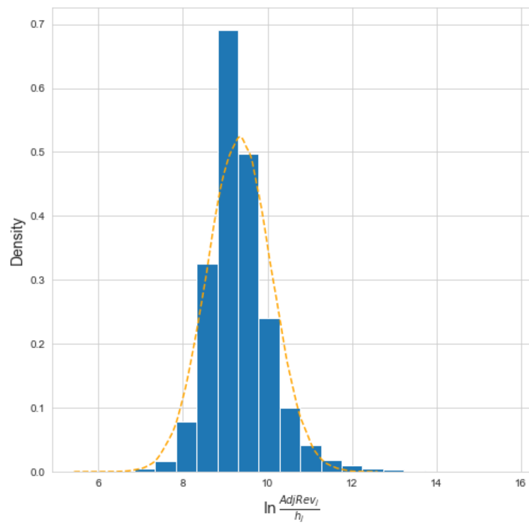
	(1)	(2)	(3)	(4)
	$\ln A_j$	$\ln l_j^{90}$	$\ln TeamSize_j$	$\ln h_{1j}$
$AfterCrisis_j$	0.075***	0.086***	−0.096***	−0.097***
	(0.022)	(0.021)	(0.024)	(0.023)
$\ln Y_j$		−0.436***	0.519***	0.696***
		(0.007)	(0.008)	(0.008)
Sample size	3689	3689	3689	2348
Adj. R^2	0.003	0.534	0.547	0.766

Note: Estimation results of Equation (14). Sample of real data includes jobs with $2005 \leq StartYear \leq 2007$ or $2010 \leq StartYear \leq 2012$. Regressions using real data control for industry and job content fixed effects. Standard errors are reported in parentheses. Statistical significance at 10%, 5%, 1% level on two-tailed tests, is indicated by *, **, ***.

Table 8: Model Validation

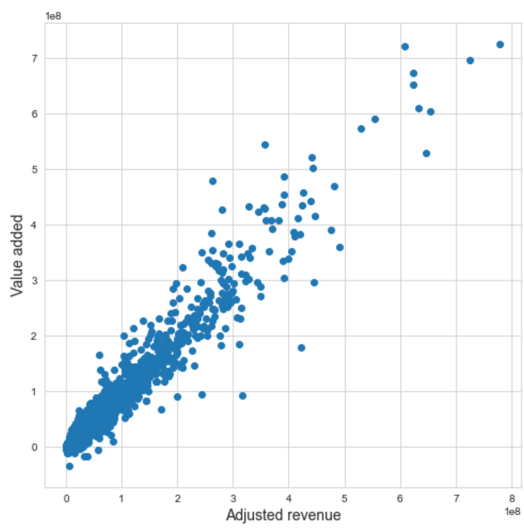
Average change in team productivity	$\frac{1}{JP}\sum_{j,p}\left(\ln A_{jp} - \ln A_j\right)$	0.076
Average labor reallocation	$\frac{1}{JP}\sum_{j,p}\left(\ln A_{jp} - \ln \overline{A}_{jp}\right)$	0.052
Average change in individual worker productivity	$\frac{1}{JP}\sum_{j,p}\left(\ln \overline{A}_{jp} - \ln A_j\right)$	0.024

Table 9: Decomposition based on nearest feasible allocation



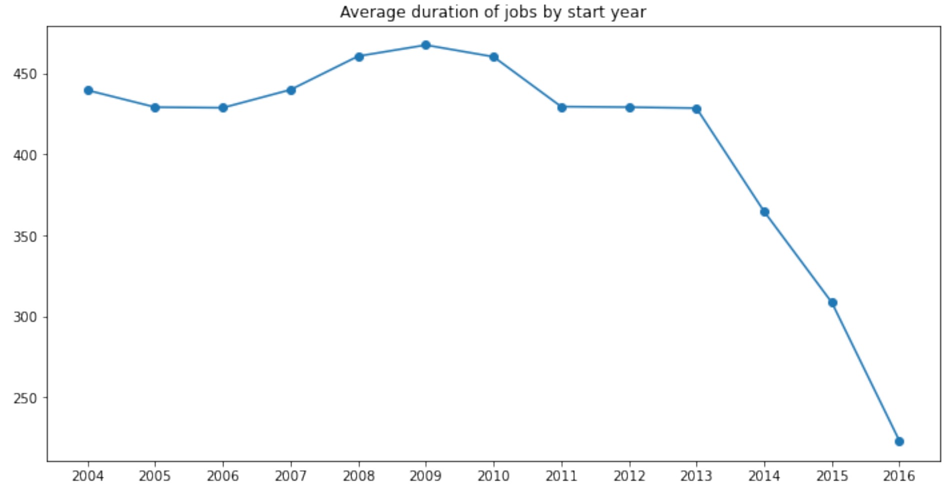
Note: Histogram of $\ln \frac{AdjRev_j}{h_j}$ overlaid with the density function (depicted with a dashed line) of a normal random variable with the same mean and standard deviation as those of the histogram.

Figure 1: Distribution of log productivity in data



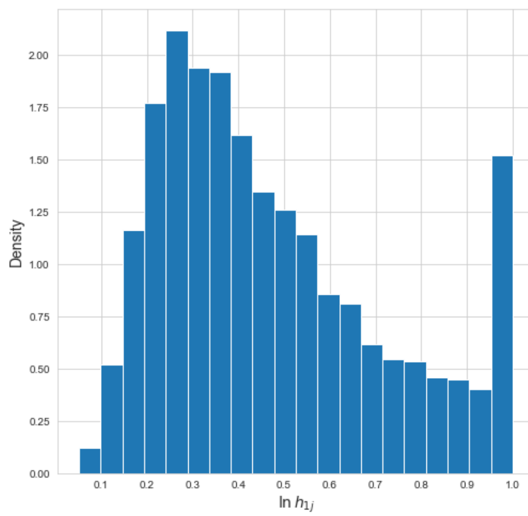
Note: Adjusted revenue for each job is plotted on the horizontal axis, and the corresponding value added measure is plotted on the vertical axis.

Figure 2: Adjusted revenue and value added



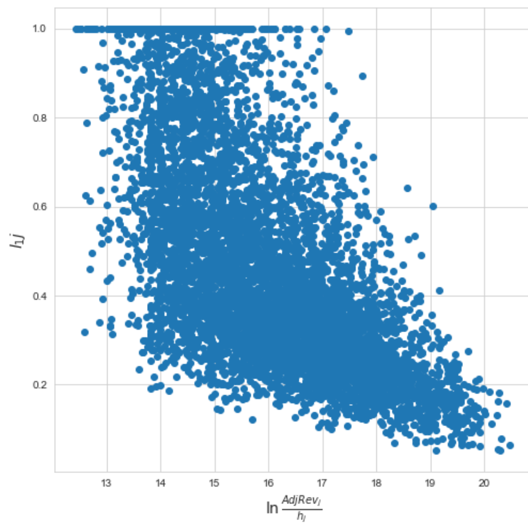
Note: Average duration (in days) of jobs that start in each year and are completed before the end of the sample period.

Figure 3: Average duration of jobs, 2004 to 2016



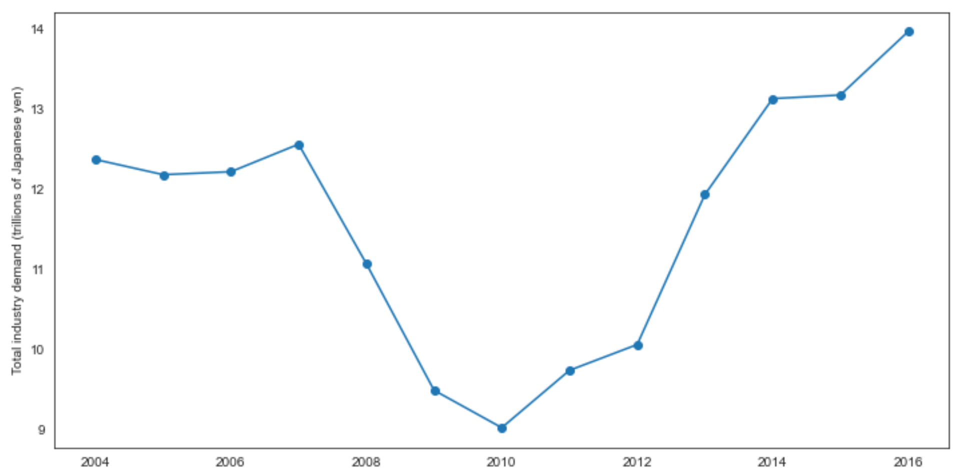
Note: Histogram of l_{1j}, or the fraction of hours contributed by the rank-1 worker.

Figure 4: Empirical distribution of the fraction hours contributed by the rank-1 worker



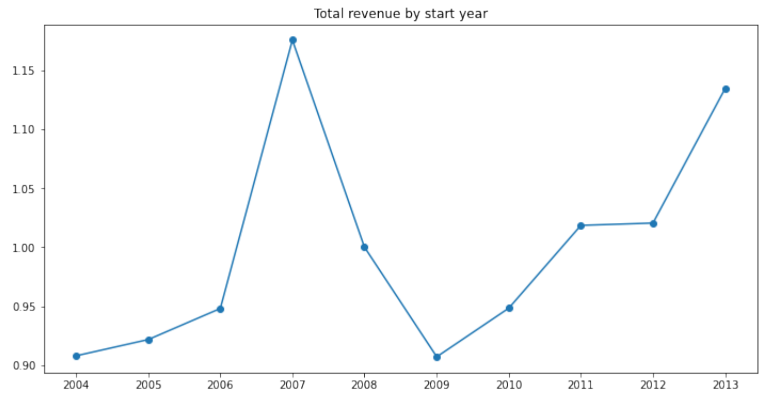
Note: Fraction of hours contributed by the rank-1 worker, l_{1j}, is on the vertical axis, and $\ln AdjRev_j$ is on the horizontal.

Figure 5: Team productivity, rank-1 worker's hours fraction, and output size



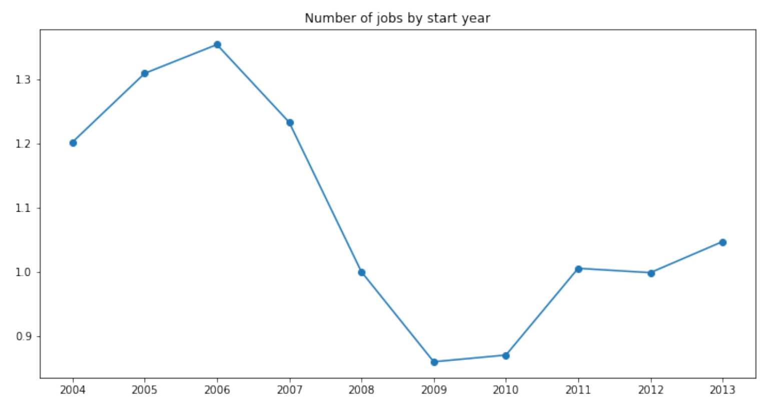
Note: Total annual amount of orders in the industry survey. Data are from the Current Survey on Orders Received for Construction, conducted by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT).

Figure 6: Industry demand, 2004 to 2016



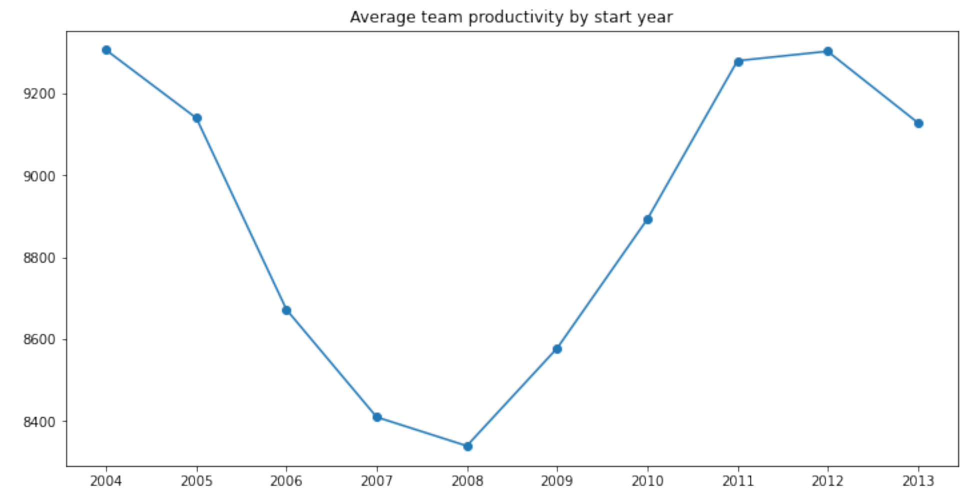
Note: Sum of the revenues from the jobs starting in each year. Year 2008 normalized to 1.

Figure 7: Total revenue by start year, 2004 to 2013



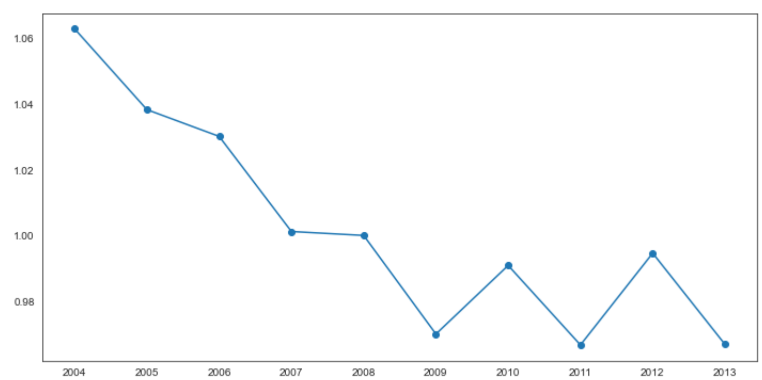
Note: Number of jobs starting in each year with revenue no smaller than one million Japanese yen. Year 2008 is normalized to 1.

Figure 8: Number of jobs by start year, 2004 to 2013



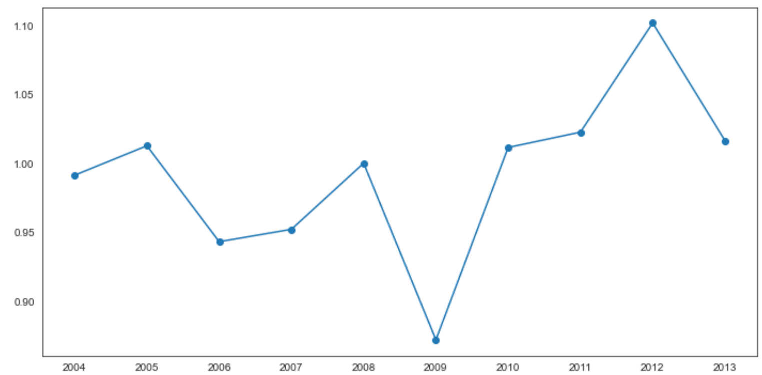
Note: For each year, the figure plots average adjusted revenue per hour, $\frac{AdjRev_j}{h_j}$, for the selected sample of jobs, with the average calculated as the ratio between the sum of adjusted revenue and the sum of total hours over all jobs starting in the same year.

Figure 9: Revenue per hour, 2004 to 2013



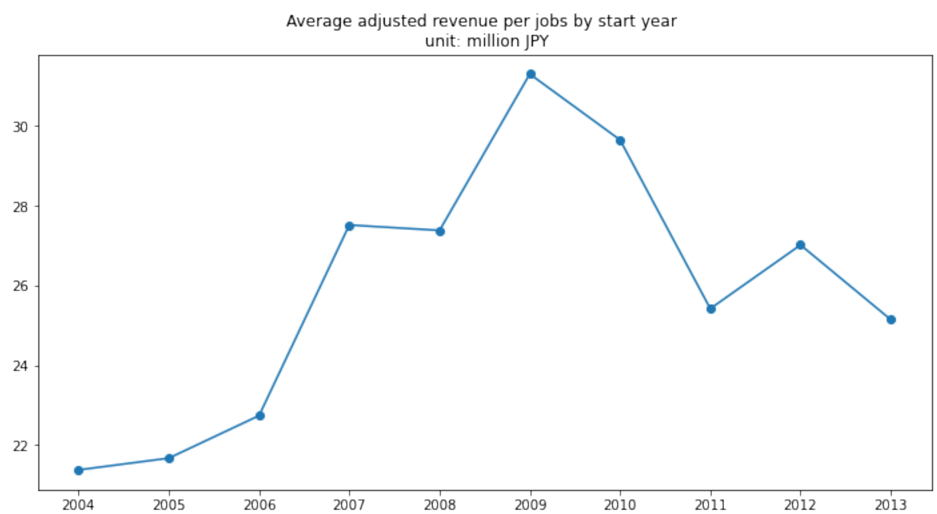
Note: Average age of rank-1 worker across jobs that start in each year (from 2004 to 2013), after controlling for client industry and job content fixed effects. Year 2008 is normalized to 1.

Figure 10: Average rank-1 worker's age, 2004 to 2013



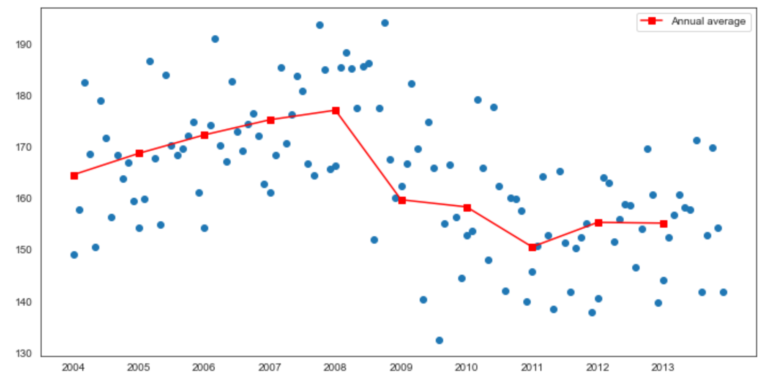
Note: Starting year fixed effects (from 2004 to 2013) after controlling for the time trend of the average rank-1 worker's age. Year 2008 is normalized to 1.

Figure 11: Revenue per hour adjusted by time trend of average rank-1 worker's age, 2004 to 2013



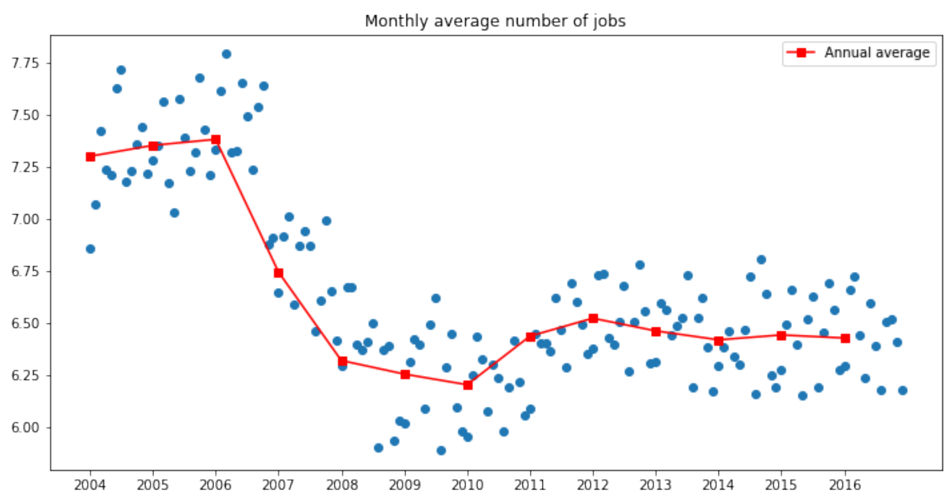
Note: Average adjusted revenue per job (2004 to 2013).

Figure 12: Adjusted revenue per job, years 2004 to 2013



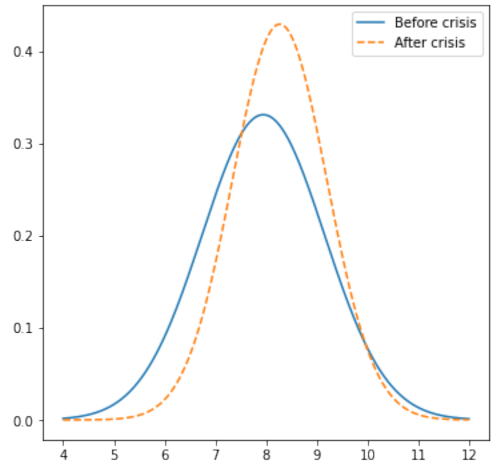
Note: Each point in the scatter represents the average monthly total working hours across workers. Only working hours on revenue-generating jobs are included. The solid line connects the square points, each of which is an annual average.

Figure 13: Average working hours, 2004 to 2013



Note: Each point in the scatter represents the average number of jobs in which each worker participates. The solid line connects the square points, each of which is an annual average.

Figure 14: Average number of jobs assigned to each worker



Note: Implied distribution of $\ln \phi_i$, before and after the crisis.

Figure 15: Calibrated density function of $\ln \phi_i$

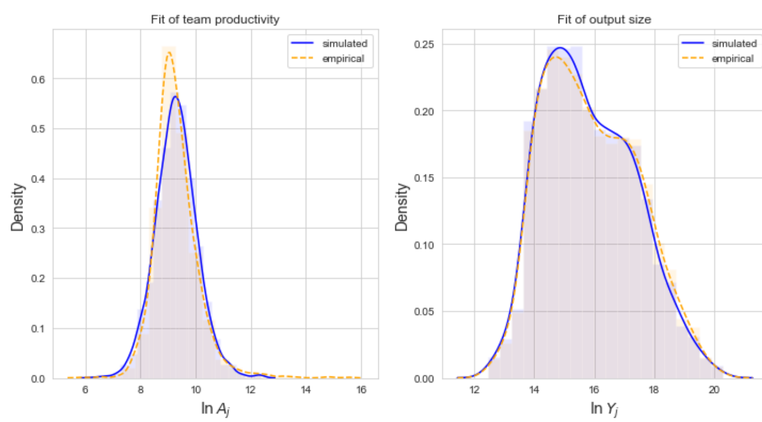


Figure 16: Fit of A_j, Y_j, before crisis

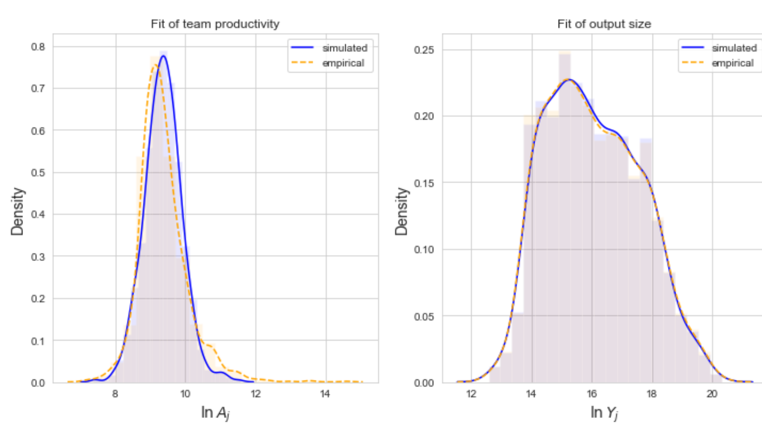


Figure 17: Fit of A_j, Y_j, after crisis

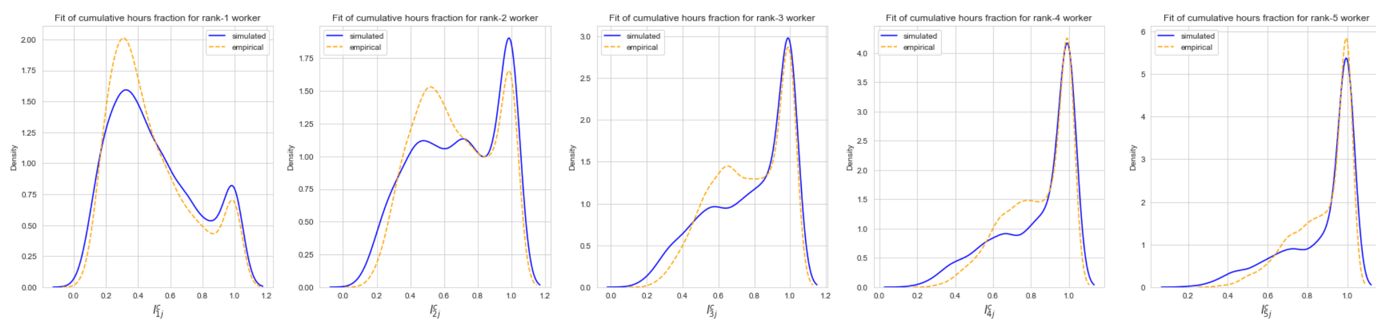


Figure 18: Fit of $l^c_{1j}, \dots, l^c_{5j}$ before crisis

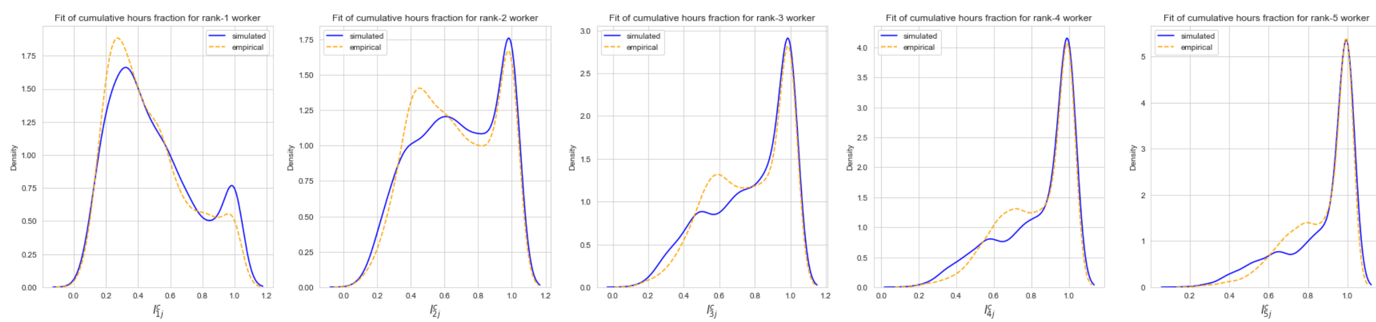


Figure 19: Fit of l^c_{1j}, ... , l^c_{5j} after crisis

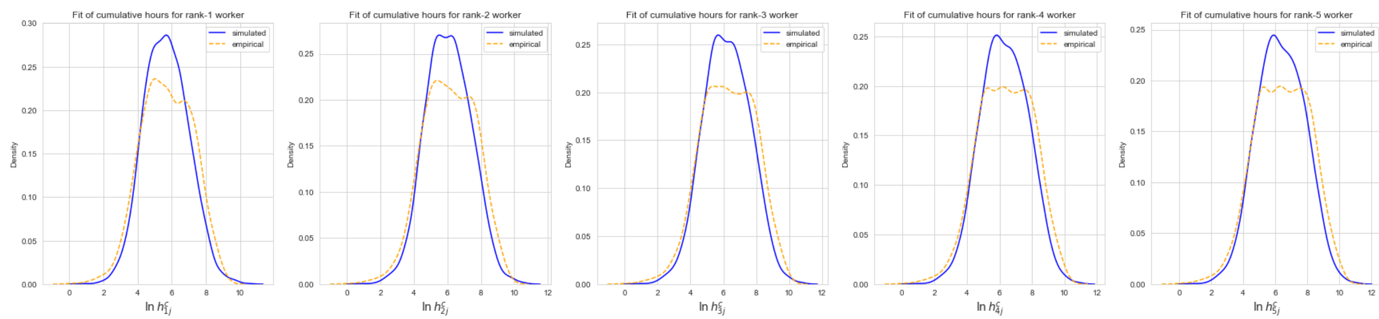


Figure 20: Fit of $\ln h^c_{1j}$, ... , $\ln h^c_{5j}$ before crisis

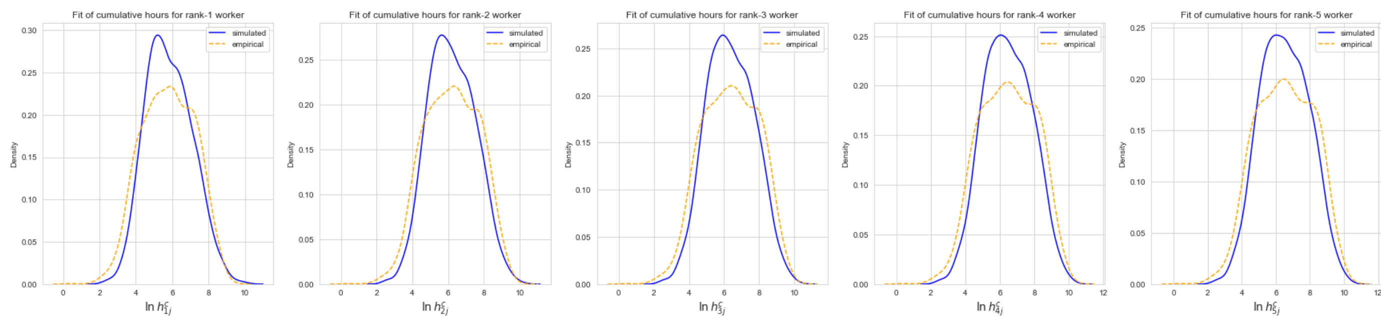


Figure 21: Fit of $\ln h^c_{1j}, \ldots, \ln h^c_{5j}$ after crisis

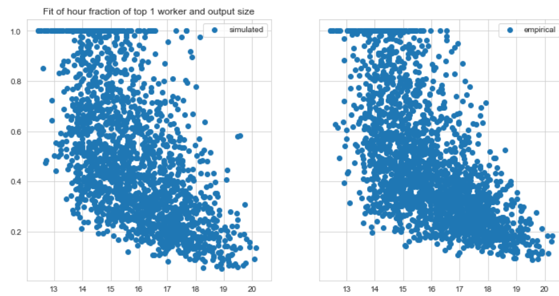


Figure 22: Fit of l_{1j} conditional on $\ln Y_j$ before crisis

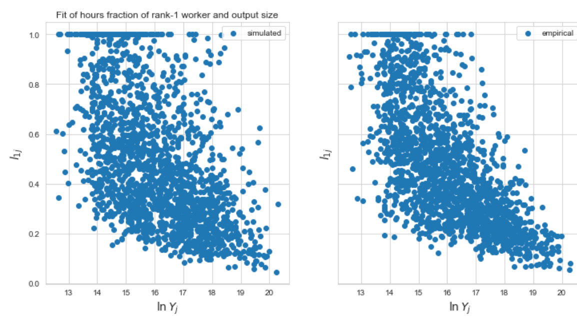


Figure 23: Fit of l_{1j} conditional on $\ln Y_j$ after crisis

Appendix

A Model solution

Given the team composition, meaning the values of ϕ_{ij} and H_{ij} for all workers on team j, the within-team labor allocation problem is

$$\min_{h_{ijs}, M_{ij}} \left(\sum_i w_{ij} M_{ij} h_{ijs} \right), \tag{19}$$

subject to

$$Y_j = \exp\left[\sum_i \frac{M_{ij}}{S} \ln\left(\phi_{ij} h_{ijs}\right) \right],$$

$$q_{ijs} = \phi_{ij} h_{ijs},$$

$$w_{ij} = c_{ij} \phi_{ij},$$

$$M_{ij} h_{ijs} \leq H_{ij},$$

$$\sum_i M_{ij} = S.$$

We first ignore workers' time constraints and take M_{ij} as given to solve for the optimal h_{ijs}. The Lagrangian is

$$\mathcal{L} = \sum_i c_{ij} \phi_{ij} M_{ij} h_{ijs} + \lambda \left(Y_j - \exp\left[\sum_i \frac{M_{ij}}{S} \ln\left(\phi_{ij} h_{ijs}\right) \right] \right)$$

The first-order condition (for task s_0 assigned to worker i_0) is:

$$\frac{\partial \mathcal{L}}{\partial h_{i_0 j s_0}} = c_{i_0 j} \phi_{i_0 j} M_{i_0 j} - \lambda \exp\left[\sum_i \frac{M_{ij}}{S} \ln\left(\phi_{ij} h_{ijs}\right) \right] \frac{M_{i_0 j}}{S} \frac{1}{h_{i_0 js}} = 0.$$

Since at the optimum $Y_j = \exp\left[\sum_i \frac{M_{ij}}{S} \ln\left(\phi_{ij} h_{ijs}\right) \right]$, we have

$$q_{ijs_0} = \frac{\lambda S^{-1} Y_j}{c_{i_0 j}},$$

so that for two tasks (s_0 and s_1) assigned to worker i_0 and i_1,

$$\frac{q_{ijs_0}}{q_{ijs_1}} = \frac{c_{i_0 j}^{-1}}{c_{i_1 j}^{-1}}.$$

Let $q_{ijs} = Bc_{ij}^{-1}$ for all s. Then

$$Y_j = \exp\left[\sum_i \frac{M_{ij}}{S} \ln\left(Bc_{ij}^{-1}\right)\right],$$

which can be rewritten as

$$B = \frac{Y_j}{\exp\left(\sum_i \frac{M_{ij}}{S} \ln c_{ij}^{-1}\right)}.$$

Then

$$q_{ijs} = \frac{c_{ij}^{-1}}{\exp\left(\sum_i \frac{M_{ij}}{S} \ln c_{ij}^{-1}\right)} Y_j,$$

and (using $q_{ijs} = \phi_{ij} h_{ijs}$)

$$h_{ijs} = \frac{c_{ij}^{-1}}{\exp\left(\sum_i \frac{M_{ij}}{S} \ln c_{ij}^{-1}\right)} \frac{Y_j}{\phi_i} = \frac{R_{ij} Y_j}{\phi_{ij}}, \tag{20}$$

where $R_{ij} = \frac{c_{ij}^{-1}}{\exp\left(\sum_i \frac{M_{ij}}{S} \ln c_{ij}^{-1}\right)}$.

Next, plugging in the optimal h_{ijs} into the objective function, total costs are equal to

$$\sum_i w_{ij} h_{ij} = SY_j \prod_i \left(c_{ij}\right)^{\frac{M_{ij}}{S}}.$$

Therefore, the optimal M_{ij} is determined by assigning as many tasks to workers with the lowest c_{ij} as possible. Under the assumption that workers with higher ϕ_{ij} have lower c_{ij}, the optimal assignment rule first exhausts the time endowment of the most productive worker before assigning the worker with the second-highest value of ϕ_{ij}, and so on, until the required output, Y_j, is achieved.

B Calibration and simulation

We begin by describing how we calibrate $(\alpha_1, \mu_H, \sigma_H^2)$ directly from the data, starting with α_1. Given that the size of the candidate pool determines the upper bound of team size, we choose a value for α_1 such that the maximum size of the candidate pool is somewhat higher than the observed largest team size. Given that the largest team sizes are similar before and after the crisis (i.e., 98 before and 97 after), we hold the size of candidate pool for the largest job constant before and after the crisis. Specifically, we choose the value of α_1 such that the job with highest output size in the sample is given 300 draws, or $\alpha_1 = \tilde{\alpha}_1 \left(Y_{max}^{-\alpha_0} \right)$,where Y_{max} is the largest revenue size in the sample, and $\tilde{\alpha}_1 = 300$. The number 300 is large enough that the subsequent team selection process could reasonably result in a team of 100, which is roughly the maximum team size observed in the data. Note that this assumption makes α_1 a function of α_0. Therefore, the value of α_1 is chosen together with α_0 in a later stage of simulation. Since we choose other parameter values to match the empirical moments, the choice of $\tilde{\alpha}_1$ does not have a real impact on the model behavior.[54]

Next, parameter values for $\left(\mu_H, \sigma_H^2\right)$ are assigned using the empirical average and standard deviation of the hours of the workers. For teams with five or fewer workers, we use the rank-1 to rank-(n_j-1) workers. For example, in a team with three workers (i.e., n_j = 3), only the hours of the two workers who contribute the most to the team's total working hours are included in the computation of the mean and standard deviation. For teams with more than five workers, we include only the rank-1 to rank-4 workers. Both μ_H and σ_H^2 are computed separately for the pre-crisis and the post-crisis samples to allow both parameters to potentially change over time. The optimal assignment rule in the model implies that all workers in the team, except possibly for the least productive one, exhaust their time endowments. Therefore, the distribution of observed hours is informative about the distribution of time endowments because the two distributions coincide for all workers (except possibly for the least productive one). But including workers with trivial hours contributions (e.g., those beyond the team's top four contributors of hours) are likely to introduce noise into the calibration. We find that our measure matches the observed hours reasonably well.

Simulation of the model requires us to specify the worker's cost per effective hour c_{ij}. From the expression for R_{ij}, i.e., $R_{ij} = \frac{c_{ij}^{-1}}{\exp\left(\sum_i \frac{M_{ij}}{S} \ln c_{ij}^{-1}\right)}$, multiplying all c_{ij} by a constant

[54] For example, if the number of draws increases, then the likelihood of taking a draw of a productive worker would be higher, but changing the parameters of the productivity distribution could result in the same effects. When we set $\tilde{\alpha}_1$ at 200 rather than 300, the results do not change qualitatively.

will not change the value of R_{ij}. Since c_{ij} affects labor allocation only through the value of R_{ij}, for the purpose of simulation, estimating the value of c_{ij} up to a proportion (equivalently, the value of $\ln c_{ij}$ up to a constant) is sufficient. We estimate the distribution of $\ln c_{ij}$ (up to a constant) using the following steps. First, using the salary data from 2012 to 2016, we estimate the wage per hour that is driven by variation in worker productivity, by estimating the following regression:

$$\ln Salary_{iy} = \beta_0 + \beta_1 \ln Tenure_{iy} + \psi_{iy}^{HL} + \psi_y^{Year} + e_{iy},$$

where $Salary_{iy}$ is the annual salary for worker i in year y, $Tenure_{iy}$ is worker $i's$ tenure in year y, ψ_{iy}^{HL} indicates individual effects for worker i's level within the firm's job hierarchy in year y, and ψ_y^{Year} indicates year effects.[55] We interpret the residual from this regression as an estimate of $\ln w_{iy}h_{iy}$, based on the idea that the remaining variation in the annual salary after controlling for tenure, workers' job rank, and year effects captures the salary variation driven by productivity differences. We take the wage distribution as time invariant because we only have salary data from 2012 to 2016. Subtracting $\ln h_{iy}$ from $\ln w_{iy}h_{iy}$ yields an estimate of the logarithmic value of the wage, $\ln w_{iy}$. After subtracting the sample median from this estimated value of $\ln w_{iy}$, we obtain a distribution of normalized logarithmic hourly wages across workers.[56]

Given the parameters of the productivity distribution, we can calculate percentiles of $\ln \phi_{ij}$, normalized by subtracting the median. Recalling that w_{ij} is monotonically increasing in ϕ_{ij}, we assign the worker at percentile q in the $\ln \phi_{ij}$ distribution with the value of $\ln w_{ij}$ at percentile q in the $\ln w_{ij}$ distribution. Finally, recalling that $\ln w_{ij} = \ln c_{ij} + \ln \phi_{ij}$, we can calculate the normalized $\ln c_{ij}$ at percentile q in the $\ln \phi_{ij}$ distribution by subtracting the value of the normalized $\ln \phi_{ij}$from the value of the normalized $\ln w_{ij}$.

The parameters $(\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0)$ are calibrated by matching the moments computed using the actual data to those computed using a simulated data set. To construct the simulated data set, we begin by assuming $S = 1$, recalling that S is the measure of tasks in a job.[57] The units of the marginal distribution of time endowments, H_{ij}, are measured in days to render hours comparable across jobs. Therefore, in the simulation, each worker's total time endowment, $\tilde{H}_{ij}$, is determined by $\tilde{H}_{ij} = H_{ij}T_j$, where H_{ij} is the daily time endowment and T_j is the job duration. The duration of each job is taken from data and

[55]The results do not change significantly in a specification that omits ψ_{iy}^{HL}.

[56]Normalization by subtracting sample mean yields a similar distribution of $\ln w_{iy}$.

[57]The choice of S does not have real effects in the model. To see why, note that if S is doubled while workers' productivities are increased by the same factor, then everything remains the same except that the measure of tasks assigned to each worker doubles.

treated as exogenous.[58] The following 3-step procedure yields values for $(\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0)$. First, for each job j in the sample, simulated values of Y_j are drawn from the real data. Second, various moments (defined below) are constructed using both the empirical and simulated data. Third, values for $(\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0)$ are chosen that minimize a distance function that is the sum of the squared deviations of the empirical moments from their corresponding simulated values.

We next describe the simulation procedure in more detail. Consider the team formation stage of the model. Given the simulated values for Y_j , the values of $\left(\mu_H, \sigma_H^2, \alpha_1\right)$, and random values for $(\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0)$, N_j draws are taken from the joint distribution of $\left(\ln \phi_{ij}, \ln H_{ij}\right)$,recalling that N_j is determined from Equation (10).[59] Then workers are assigned to tasks as described in section 4.1. As described in the model, the labor assignment of each simulated job j takes Y_j as given. Working hours, team productivity, and the total hours contributed by the workers (by rank) are calculated according to equations (6), (9), and (7).

Calculating the optimal hours for each task, given a certain worker and job, requires knowledge of the optimal task assignment M_{ij}. In general, M_{ij} must be solved for numerically using the following two-step procedure. The first step computes the realized team size. For every integer value of n_j^{temp} from 1 to N_j, we test whether a team composed of the most productive n_j^{temp} workers can complete the job. We do so by assuming that all workers exhaust their time endowments, which permits the following closed-form solution for M_{ij} :

$$\frac{M_{ij}}{S} = \frac{c_{ij}\phi_{ij}H_{ij}}{\left(\sum_{i=1}^{n_j} c_{ij}\phi_{ij}H_{ij}\right)}. \tag{21}$$

If the total measure of tasks allocated under this assumption exceeds S, then the team of n_j^{temp} workers can complete the job. If n_j workers can complete the job but $n_j - 1$ cannot, the realized team size is n_j, and the loop for finding the realized team size stops. If a team of all N_j workers cannot complete the job, the job is regarded as failed and is dropped

[58]The contracted job duration is determined by a deadline that is stated within the contract that is signed at the start of the job. Although the observed job duration may differ from the contracted job duration due to unexpected events, such events are uncommon and should be independent across projects in the sample.

[59]Drawing from the joint distribution of $\ln \phi_{ij}$ and $\ln H_{ij}$ involves first taking draws a_{ij} and b_{ij} from the standard normal distribution and then using the following matrix multiplication:

$$\begin{pmatrix} \ln \phi_{ij} - \mu_\phi \\ \ln H_{ij} - \mu_H \end{pmatrix} = \begin{pmatrix} \sigma_\phi \sqrt{1-\rho_{\phi H}^2} & \sigma_\phi \rho_{\phi H} \\ 0 & \sigma_H \end{pmatrix} \begin{pmatrix} a_{ij} \\ b_{ij} \end{pmatrix}.$$

It is easy to verify that the resulting random variables have the desired joint distribution. The draws of a_{ij} and b_{ij} are fixed throughout the simulation process.

from the sample.[60] The second step solves the optimal task assignment across the n_j workers. Since the most productive $n_j - 1$ workers exhaust their time endowment, we have for $i = 1, 2, ...n_j - 1$,

$$M_{ij}h_{ijs} = M_{ij}\frac{c_{ij}^{-1}}{\exp\left(\sum_i \frac{M_{ij}}{S}\ln c_{ij}^{-1}\right)}\frac{Y_j}{\phi_{ij}} = H_{ij}, \tag{22}$$

and

$$\sum_{i=1}^{n_j} M_{ij} = S. \tag{23}$$

Let $D_j = \exp\left(\sum_i \frac{M_{ij}}{S}\ln c_{ij}^{-1}\right)$. From the above equations, we can express M_{ij} as a function of D_j for all workers. Plugging the expressions for M_{ij} into the definition of D_j reveals that for job j, D_j satisfies

$$\begin{aligned} \ln D_j &= S^{-1}Y_j^{-\eta}D_j\sum_{i=1}^{n_j-1} c_{ij}\phi_{ij}H_{ij}\ln c_{ij}^{-1} \\ &\quad + \left(1 - S^{-1}Y_j^{-\eta}D_j\sum_{i=1}^{n_j-1} c_{ij}\phi_{ij}H_{ij}\right)\ln c_{n_j j}^{-1}, \end{aligned} \tag{24}$$

which is a nonlinear equation for a single variable. By solving this equation for D_j numerically, we can then solve the optimal task assignment M_{ij} for all workers.

Values for $\left(\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0\right)$ are chosen by matching the moments calculated using the simulated sample and the corresponding empirical moments. The procedure is applied twice to obtain two sets of values (pre-crisis and post-crisis) for $\left(\mu_\phi, \sigma_\phi^2, \rho_{\phi H}, \alpha_0\right)$. The targeted moments are:

- distribution of job-level productivity. The model predicts that, conditional on Y_j, job-level productivity is determined by the average of worker-level productivity. Therefore, the moments of the job-level productivity distribution help to identify μ_ϕ and σ_ϕ^2.

- distribution of output size, Y_j. As shown in equations (6) and (9), both the time required to complete one task and the job-level productivity are increasing in Y_j. Therefore, the observed distribution of Y_j puts restriction on the model's parameters.[61]

[60] Later, when discussing the target moments, we explain how we minimize the impact of job failures.

[61] One complication is that time constraints create the possibility that the assigned workers cannot com-

- distributions of the fraction of the team's hours contributed by its workers (by rank), cumulatively, up to the rank-5 worker.[62] Fractions of hours that are contributed (by rank) are informative about the correlation parameter $\rho_{\phi H}$. To see this, observe that given the other parameters, if $\rho_{\phi H} > 0$, more productive workers tend to have greater time endowments. Therefore, hours will be more concentrated compared to the case when $\rho_{\phi H} < 0$. The fraction of hours contributed (by rank) is targeted instead of team size because a complete profile of that fraction is sufficient to calculate team size. For example, the fractions in a two-person team could be split in an infinite number of ways.

More specifically, we solve the following optimization problem numerically:

$$\min_{\alpha_0, \mu_\phi, \sigma_\phi^2, \rho_{\phi H}} \frac{1}{3} \left(Err_A + Err_Y + Err_l \right),$$

where Err_A is the error, or the distance between the empirical and simulated moments, for moments related to team productivity, Err_Y is the error for moments related to output size, and Err_l is the error for moments related to hour fractions. Each of the three error terms is calculated as the mean squared errors between a group of empirical and simulated moments. To calculate Err_A, we include the following moments of the logarithmic value of team productivity distribution: the mean of the distribution, the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles, the ratio between the 90^{th} percentile and the 10^{th} percentile, and the ratio between the 75^{th} percentile and the 25^{th} percentile. To calculate Err_Y, we include the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of the output size distribution. And to calculate Err_l, we include the 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of the distribution of rank-k cumulative fraction of hours contributed, for k equal to 1, 2, 3, 4, and 5.[63]

C Algorithm to find the nearest feasible allocation

We use the following algorithm to identify the nearest feasible labor allocation in the counterfactual job:

1. For the set of workers who are common to both the actual and counterfactual jobs,

plete the job. The simulation drops such failed jobs from the sample. As a result, the simulated distribution of Y_j can differ from the empirical distribution of Y_j. Including the empirical distribution of Y_j in the target moments minimizes this effect.

[62] The top five workers account for most of the team's hours (about 88.3%, on average).

[63] The Basin-hopping algorithm (Wales and Doye (1997)) is applied to avoid having the optimization routine trapped at a local minimum.

namely for $i = 1, 2, 3, ..., \min\left(n_j, n_{jp}\right)$, set $M_{ij} = M_{ijp}$. If the actual team is larger than the counterfactual team (i.e., if $n_j > n_{jp}$) some measure of tasks remains unallocated in the counterfactual job. To minimize the sum of squared errors, evenly allocate the remaining measure of tasks to each worker in the counterfactual team even if doing so would violate time endowment constraints. If the actual team is smaller than the counterfactual team (i.e., if $n_j < n_{jp}$) there will be at least one worker who is initially allocated zero tasks in the counterfactual job.

2. Check the time constraints for all workers. If the time constraints are all satisfied, stop.

3. If the time constraints are violated for some workers, set the measure of tasks such that time constraints are exactly satisfied for those workers.

4. From step 3, due to the violation of time constraints, a measure of tasks remain to be allocated. To minimize the sum of squared errors, evenly allocate these tasks across the other workers whose time constraints are not violated.

5. Check the time constraints for all workers. If the time constraints are all satisfied, stop. If not, repeat from step 3.

Note that the optimal hours for each worker and task, h_{ijs}, depend on the allocation of tasks across the other team members, through the effect of the denominator in R_{ij}, or $\exp\left(\sum_i \frac{\overline{M}_{ij}}{S} \ln c_{ij}^{-1}\right)$. As a result, some workers might have their time constraint satisfied in step 3 and then later violated in step 4 upon allocation of the remaining measure of tasks. However, this effect will be smaller as iterations evolve because the resulting allocation will be closer and closer to a feasible allocation. The existence of a feasible allocation is guaranteed because we drop observations in which the team is unable to complete the job even by exhausting all team members' time endowments (i.e., job failure). In the parameter calibration described in Appendix B, we add constraints so that the risk of job failure is not substantial, and in fact it occurs for less than 1% of all jobs in this counterfactual exercise. We allow time constraints to be satisfied with an error of 10^{-5}, and convergence is always achieved within 50 iterations.